2020

Numerical simulation and data-driven analysis of infrared detector performance

BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Dissertation

NUMERICAL SIMULATION AND DATA-DRIVEN ANALYSIS OF INFRARED DETECTOR PERFORMANCE

by

ANDREU GLASMANN

B.S., Davidson College, 2014
M. Eng. Boston University, 2015

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2020

Approved by

First Reader

Enrico Bellotti, PhD
Professor of Electrical and Computer Engineering
Professor of Materials Science and Engineering

Second Reader

Sahar Sharifzadeh, PhD
Assistant Professor of Electrical and Computer Engineering
Assistant Professor of Materials Science and Engineering

Third Reader

Roberto Paiella, PhD
Professor of Electrical and Computer Engineering
Professor of Materials Science and Engineering

Fourth Reader

Marion Reine, PhD
Consultant on Infrared Detectors

There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after. J. R. R. Tolkien

Acknowledgments

This work was financially supported by the Microsystems Technology Office at the Defense Advanced Research Projects Agency, a Department of Defence ManTech program, the Boston University Ignition program, and the U.S. Army Research Laboratory's Center for Semiconductor Modeling. Computational resources were provided by U.S. Army Research Office Defense University Research Instrumentation Program awards.

I would like to extend my sincerest gratitude to my family, friends, and colleagues whose support made this work possible. My fiancé, Elizabeth, deserves special recognition for enduring countless rants and long evenings, during an especially difficult climate. I'd like to also express thanks to my parents, Ron and Beth, for their encouragement and support for my continuing education.

I am extremely grateful for the conversations and feedback from members of the Computational Electronics Group at BU, both past and present. Of special note are the previous members, Drs. B. Pinkie, J. Schuster, D. D'Orsogna, and A. Wichman, whose work paved the platform that this research built upon. I'd like to extend a personal thanks to the committee for their time and feedback throughout my time at BU, Drs. R. Paiella, S. Sharifzadeh, and M. Reine. I would like to offer my thanks to Drs. M. Nguyen, P. Wijewarnasuriya, A. D'Souza, P. Lamarre, P. Mak, V. Kleptsyn, and R. DeWames for many fruitful discussions during the various stages of my Ph. D. A special thanks is given to J. Goebel, A. McFarland, and others at ENGIT and BU IS&T for putting up with a constant stream of IT-related questions while administrating our computational resources. Finally, I want to express my gratitude to I. Prigozhin, B. Appleton, and T. Hubbard for the many hours we spent working together on the DCJ program.

Last, I offer my deepest appreciation to my advisor, Prof. E. Bellotti, for accepting

me into his research group, and providing me the resources, and time, to continue to improve my skills and explore my ideas.

Andreu L. Glasmann

NUMERICAL SIMULATION AND DATA-DRIVEN ANALYSIS OF INFRARED DETECTOR PERFORMANCE

ANDREU GLASMANN

Boston University, College of Engineering, 2020

Major Professor: Enrico Bellotti, PhD
Professor of Electrical and Computer Engineering
Professor of Materials Science and Engineering

ABSTRACT

Infrared detectors are a critical technology frequently used in commercial, military, and scientific settings. Research and development of modern infrared detectors is driven by finding new ways to reduce the system's size, weight, power, and cost while adding new functionalities without lowering performance. Physics-based numerical models can be instrumental in lowering the cost of developing such advances. This dissertation presents three main contributions to further the predictive modeling of infrared photodetectors.

First, motivated by recent demonstrations of small-pitch focal plane arrays for infrared imaging—5 µm for SWIR and 10 to 15 µm for MWIR/LWIR—we use physics-based numerical simulations to assess the implications on dark current, quantum efficiency, specific detectivity, and modulation transfer function in SWIR $In_{0.53}Ga_{0.47}As$ FPAs. From the results, we propose a new pixel sub-architecture aimed toward lowering dark current and improving MTF.

Second, we present a methodology for simulating the capacitance-voltage (C-V)

characteristics of nBn photodetectors. For junction-based semiconductor devices, C-V profiling is a common technique for non-destructively characterizing semiconductor layers in metal-insulator-semiconductor devices by using well-established analytical relations. However, this type of analysis cannot be directly applied to the barrier detector's unique architecture and the formalism must be modified. To this end, we present a modified analytical formalism based on metal-oxide-semiconductor theory, and a methodology using the drift-diffusion method; both are used to explore the role of the device architectural properties on determining the C-V characteristics.

Last, we present several cases of applying neural networks to nBn photodetector figures of merit. We use artificial neural networks as surrogate models for the capacitance-voltage, current-voltage, and quantum efficiency to explore the multi-dimensional parameter space to assess parameter-performance correlations and determine the global role of each feature in shaping each characteristic without the need for additional simulations. Moreover, using inspiration from image recognition, we demonstrate that a convolutional neural network can be trained to analyze a C-V characteristic to yield more information about a device than what would be possible from a conventional analysis.

Contents

List of Tables

List of Figures

List of Abbreviations

1D	.....	One-Dimensional
2D	.....	Two-Dimensional
3D	.....	Three-Dimensional
AL	.....	Absorber Layer
AlAsSb	.....	Aluminum Arsenide Antimonide
ALD	.....	Atomic Layer Deposition
ANN	.....	Artificial Neural Network
BL	.....	Barrier Layer
C-V	.....	Capacitance-Voltage
CdTe	.....	Cadmium Telluride
CL	.....	Contact Layer
CNN	.....	Convolutional Neural Network
DCJ	.....	Diffusion Control Junction
DD	.....	Drift-Diffusion
DLPH	.....	Double Layer Planar Heterojunction
D^*	.....	Specific Detectivity
ELU	.....	Exponential Linear Unit
FDTD	.....	Finite-Difference Time-Domain
FPA	.....	Focal Plane Array
GaSb	.....	Gallium Antimonide
GB	.....	Gaussian Beam
GeSn	.....	Germanium Tin
GR	.....	Generation-Recombination
HgCdTe	.....	Mercury Cadmium Telluride
InAsSb	.....	Indium Arsenide Antimonide
InGaAs	.....	Indium Gallium Arsenide
InP	.....	Indium Phosphide
IR	.....	Infrared
LPE	.....	Liquid phase epitaxy
LWIR	.....	Long Wave Infrared
MBE	.....	Molecular Beam Epitaxy
MCT	.....	Mercury Cadmium Telluride
ML	.....	Machine Learning
MOCVD	.....	Metal-Organic Chemical Vapor Deposition

MOS	.....	Metal-Oxide-Semiconductor
MTF	.....	Modulation Transfer Function
MWIR	.....	Mid Wave Infrared
nBn	.....	Barrier Detector with n-type absorber and contact layers
PECVD	.....	Plasma-Enhanced Chemical Vapor Deposition
PSF	.....	Point Spread Function
QE	.....	Quantum Efficiency
ROIC	.....	Readout Integrated Circuit
SiGe	.....	Silicon Germanium
SiN	.....	Silicon Nitride
SiO_2	.....	Silicon Dioxide
SRH	.....	Shockley-Read-Hall
SS	.....	Spot-Scan
SWIR	.....	Short Wave Infrared
VLWIR	.....	Very Long Wave Infrared

Chapter 1

Introduction

1.1 An introduction to infrared photodetectors and imaging systems

Infrared (IR) radiation was first discovered by Frederick William Herschel in 1800. Using a prism to spatially separate the spectrum of sunlight, he found that the region just beyond the visibly red portion was able to heat a set of thermometers (Herschel, 1800). Though rudimentary, this also constituted the first infrared detector. Progress on developing infrared sensors progressed steadily through the following years. The 1800's focused on thermal detectors that exploit the thermoelectric property of two dissimilar metals at different temperatures as a method of detecting heat, with the first thermopile being invented in 1830 (Rogalski, 2012). The first photon detectors using photoconductivity were developed in 1930. World War I and II saw the development of photocathodes and lead sulfide based photoconductors. Small bandgap semiconductor alloys using III-V and II-VI materials were first synthesized in the 1950's and 1960's. First generation infrared imaging systems were linear scanning arrays where the scene was imaged one line at a time by mechanically rotating a mirror and measuring the signal from each pixel (Kinch, 2000, Rogalski, 2010). The major disadvantages of these linear systems was a required one-to-one electrical connection to each individual pixel, limiting the achievable resolution of the sensor and inherently low framerates. Second generation systems introduced two-dimensional staring arrays where the device is hybridized to a separate readout integrated circuit

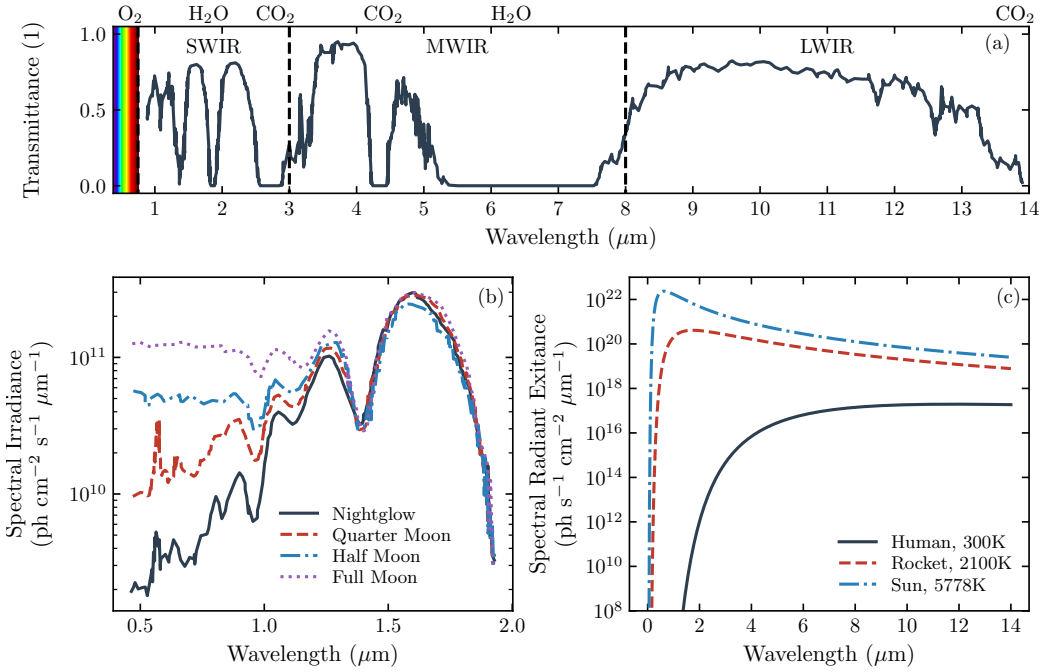

Figure 1·1: (a) Infrared transmittance of the Earth's atmosphere (Vatsia, 1972). (b) Spectral irradiance at night for different phases of the moon in the shortwave infrared (DeWames et al., 2015). (c) Blackbody spectrum for objects of different temperatures.

(ROIC) (Rogalski et al., 2009). The separate ROIC enables signal multiplexing, reducing the number of electrical connections to the number of rows and columns in the array, greatly increasing the possible number of pixels. ROIC technology offers other benefits as well, such as various image processing techniques such as anti-blooming and preamplifiers. Modern sensors, known as third generation infrared systems, further improve upon previous iterations by adding multi-spectral capabilities, achieving higher sensitivities, higher framerates, smaller pixels, larger image formats, and have steadily introduced new materials and pixel architectures (Rogalski et al., 2009, Horn et al., 2003).

Infrared detectors are often categorized by their intended spectral band of operation. Shown in Fig. 1·1(a), the short wave infrared (SWIR) spans 0.9–2.5 µm the mid wave infrared (MWIR) covers 3–5 µm long wave infrared (LWIR) includes

8–14 µm and the very long wave infrared (VLWIR) extends past 14 µm. These distinctions come from the transmission windows for terrestrial applications, where certain wavelengths of electromagnetic radiation are absorbed or transmitted by the molecules present in the Earth's atmosphere. All objects emit radiation based on its temperature—higher temperatures lead to higher emissivities at higher energy wavelengths of electromagnetic radiation. Planck's law describes the spectral content of the radiation as (Rogalski, 2010)

$$M(\lambda, T) = \frac{2\pi c}{\lambda^4} \left[\exp \frac{hc}{\lambda k T} - 1 \right]^{-1}, \tag{1.1}$$

where M, h, c, λ, k, and T are the blackbody spectral radiant exitance, Planck's constant, the speed of light, wavelength, Boltzmann constant, and temperature respectively. Shown in Fig. 1·1(c), higher temperature objects are more emissive at higher energy wavelengths. For example, the sun heavily emits in the visible range, a jet plume peaks in the shortwave infrared, and the human body modeled as an ideal blackbody source would emit in the longer infrared wavelengths. Accordingly, MWIR and LWIR detectors are typically developed for surveillance systems to detect human or jet signatures, as well as being less prone to atmospheric scattering than visible light. Detectors for night vision applications utilize the SWIR band, and rely heavily on external illumination sources (Dhar et al., 2013); in general SWIR is photon-starved with the only sources of illumination from background nightglow and reflected sunlight from the moon's surface, as shown in Fig. 1·1(b).

In basic terms, a detector absorbs incident radiation and produces a measurable change in a physical quantity. There are two types of infrared detectors: thermal and photon. Thermal detectors, such as thermopiles, pyroelectrics, and bolometers use the resulting change in the material's temperature to measure a change in voltage, polarization, and electrical resistance respectively. These types of detectors are well-

suited to low-cost applications, due to their slow response and sensitivity, with their primary advantage being the ability to operate at room temperature (Rogalski, 2010).

In this disseration, we focus on a second class of detector: photon detectors. Photon detectors use the electro-optical response of a material to sense a change in the material's electrical resistance. An incident photon provides enough energy for a bound valence electron to be excited into the conduction band, creating an electron-hole pair. These optically generated charge carriers are collected by a contact and sensed as an electrical signal in an external circuit. There are four broad classes of infrared detector materials commonly used today: extrinsic single element crystals, chemical compounds, alloys, and lattice-engineered structures (Vincent et al., 2015). Extrinsic detectors use a single element crystal, such as silicon or germanium, and introduce impurities that have defect states within the material's bandgap. The difference between the defect level and the conduction band must be equal to or less than the energy of the incident photons in order to generate carriers. Detectors based on chemical compounds, such as InAs or InSb, and alloys, like $In_{1-x}Ga_xAs$ or $Hg_{1-x}Cd_xTe$, rely on the material having a direct bandgap of appropriate energy. These types of detectors are commonly used as photovoltaics with *pn*- or *pin*-junctions. The lattice-engineered materials exploit the band alignment of two dissimilar semiconducting materials to form quantum wells or strained layer superlattices that exploit the quantum mechanical properties of electrons and holes to operate. In this dissertation we will only discuss photodetectors based on semiconducting alloys.

1.2 Infrared detector materials

Most high performance infrared imaging sensors based on semiconducting compounds or alloys are composed of II-VI or III-V materials. Some single crystal binary compounds, such as InSb, have been widely adopted for use in infrared photodetectors

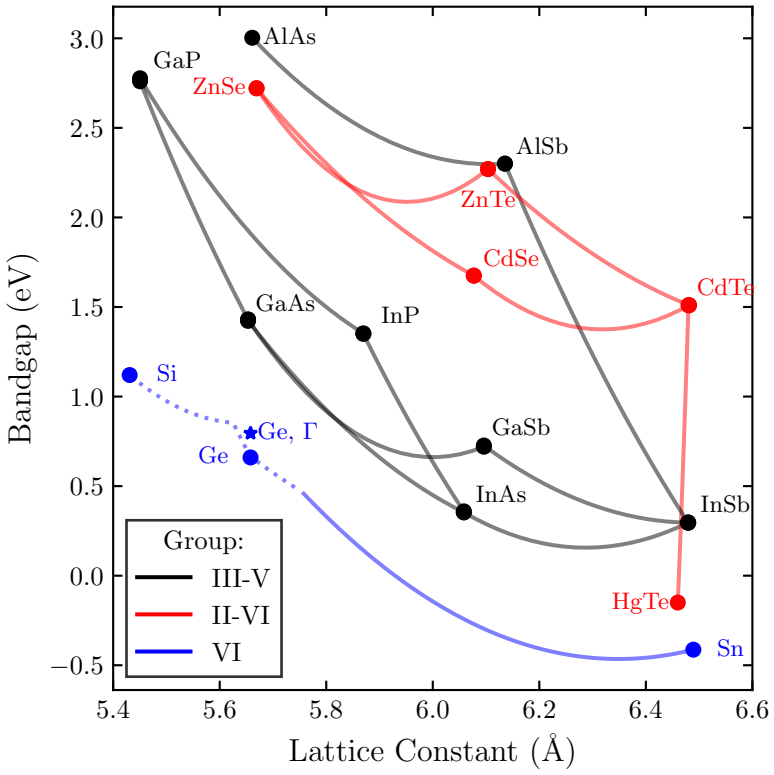

Figure 1·2: Lowest bandgap as a function of lattice constant for various zinc blende group II-VI, III-V, and IV compounds and a few of their corresponding alloys. Solid lines are direct, dotted are indirect. Data collected from literature (Kasap, 2006, Adachi, 2009, Vurgaftman et al., 2001).

due to the ability of growing high quality material in bulk with large wafer diameters. One limitation is the cutoff wavelength is fixed depending on the material's bandgap, restricting the spectral windows and applications that the material can be used for. One workaround is to alloy these compounds with an additional element, or elements, to form ternary or quaternary alloys. These alloys are usually deposited as thin films on compatible substrates through epitaxial growth techniques, such as liquid phase epitaxy (LPE), metalorganic chemical vapor deposition (MOCVD), or molecular beam epitaxy (MBE). More information on the various growth techniques can be found in other texts (Kasap, 2006). A few of the most important materials and their ternary alloys for infrared optoelectronic devices are shown in Fig. 1·2. Of special note are $In_{1-x}Ga_xAs$, $Hg_{1-x}Cd_xTe$, and $InAs_{1-x}Sb_x$. $In_{1-x}Ga_xAs$ is lattice-matched to InP with an indium content of 53%, permitting high quality material growth on InP with a 1.7 µm cutoff wavelength for SWIR applications. $InAs_{1-x}Sb_x$ with small antimony concentrations provides cutoff wavelengths between 4 and 5 µm on GaSb substrates. $Hg_{1-x}Cd_xTe$ is undeniably the most noteworthy material for infrared photodetectors as its bandgap can be tailored over the entire infrared spectrum with minimal change in lattice constant. However, challenges in material growth and device fabrication continue to plague MCT technology, leading to low device yield; thus new materials and devices are continually being funded for investigation (Capper et al., 2011, Rogalski, 2012). There have also been research in using Si_xGe_{1-x} alloys for SWIR and Ge_xSn_{1-x} alloys for MW/LWIR photodetectors. The former offers promise of monolithic integration into existing silicon technologies without separate hybridization, and enhanced absorption strength when compared to plain silicon. The latter shows promise due to the bandgap extending into the longer wave infrared wavelengths to possibly compete with MCT or antimony-based detectors. Development of new alloys for optoelectronic devices is still an active area of research today.

1.3 Infrared detector figures of merit

This dissertation is concerned with modeling infrared photodetector performance. In this section we introduce the figures of merit used when comparing different detectors. The basic operation of a photovoltaic device is shown in Fig. 1·3. Generally, infrared detectors are operated in photoconductive mode, with a small reverse bias applied to the anode. There is a small leakage current under dark conditions, referred to as dark current. Once the device is illuminated the current-voltage characteristics change such that a larger photocurrent is observed under reverse bias.[1]

1.3.1 Dark current

Dark current is one of the most important quantities when discussing infrared photodetectors. There is a current that flows when a voltage is applied to a semiconductor device. The magnitude of this current in the absence of any optical input is the dark current, and sets a limit on the minimum detectable optical signal. Infrared detectors based on small bandgap alloys or compounds inherently have higher amounts of dark current as there are more carriers available for conduction. As such, these types of detectors, especially for MWIR and LWIR, are cooled under operation to reduce the noise level until they are background-limited (Vincent et al., 2015, Rogalski, 2010).

Once a voltage is applied, or light is shined on the material, the system is no longer in equilibrium. Figure 1·3 summarizes the physical processes in a semiconducting material that try to bring the device back to equilibrium by generating or recombining carriers. Several of these processes are intrinsic to the material. Using the thermal energy of the material, a bound valence electron can gain enough energy to radiatively excite into the conduction band of the solid as shown in Fig. 1·3(a), leaving behind a hole in the valence band. There are also multi-carrier interactions, referred to as Auger

[1]The optical signal must be great enough for the current to exceed the background noise either from dark current or background radiation.

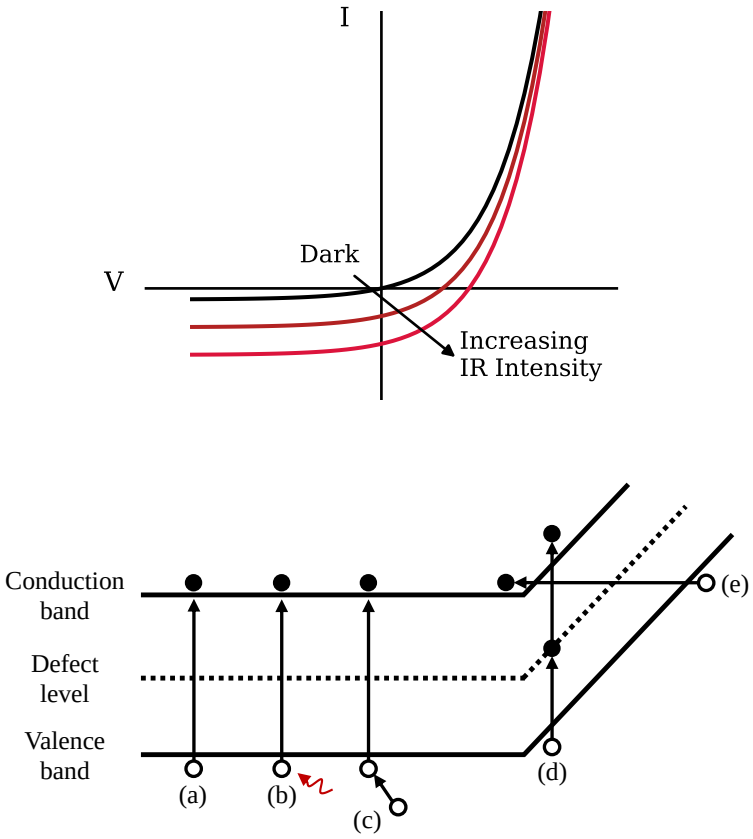

Figure 1·3: (Top) Typical photodiode current-voltage characteristics under illumination. (Bottom) Various processes for generating charge carriers in a semiconductor. (a) A radiative process where a bound valence electron is promoted to the conduction band by using the material's thermal energy, creating an electron-hole pair. (b) A photon is absorbed to provide the energy to generate free carriers. (c) A multi-particle process where one particle provides the energy create an electron-hole pair. (d) Electron-hole pairs are created through an intermediate defect state within the bandgap. (e) A bound electron in the valence band tunnels through the energy barrier to a vacant state in the conduction band.

generation-recombination, where an energetic carrier provides enough energy to the lattice to generate an electron-hole pair. In general, the best dark current performance is achieved in materials that have few defects, and are limited by the radiative and Auger mechanisms. These detectors are often referred to as being diffusion-limited, as the only source of dark current is due to these intrinsic generation mechanisms, where the generated carriers diffuse through the material and are eventually collected by a contact and sensed in an external circuit. We will discuss diffusion current in detail in subsequent chapters.

The primary source of excess dark current in photodiodes is due to defects. Defects can manifest as dislocations or point defects from contaminants and often degrade the electrical properties of the device. If the defect introduces energy states within the bandgap of the material, carriers can be more easily generated by using these intermediate states as shown in Fig. 1·3(d). This process is known as Shockley-Read-Hall (SRH) generation, and the resultant leakage current is often referred to as generation-recombination (GR) current. SRH is dominant in depleted regions of the device (Muller and Kamins, 2003, Choo, 1968). Therefore, when material quality is known to be poor and GR is the dominant source of dark current, approaches must be taken to reduce the amount of depleted material. Note, SRH defect states are not limited to the bulk material. Most exposed semiconductor surfaces have high trap densities from dangling atomic bonds that lead to excess surface leakage currents that behave similarly to bulk GR current. As such, a critical step in photodiode fabrication is properly passivating any exposed semiconducting surfaces. Silicon-based technologies have an advantage of a high quality native oxide, SiO_2. Unfortunately, most IR materials do not share this property, and require a separate passivating material, typically deposited through plasma enhanced chemical vapor deposition (PECVD), or atomic layer deposition (ALD). SiN and SiO_2 are frequently used for III-V materials,

and CdTe is used for $Hg_{1-x}Cd_xTe$.

Lastly, another important source of dark current in small bandgap materials is due to tunneling. At large biases and strong fields carriers can tunnel through the energy barrier as shown in Fig. 1·3(e), where an electron in an occupied valence band state tunnels to a vacant conduction state leading to additional dark current. Tunneling currents can be due to a direct band-to-band process or due to a defect-assisted process where an intermediate state aids the transition in trap-assisted-tunneling (TAT) (Sze and Ng, 2006, Rogalski, 2010).

1.3.2 Quantum efficiency

Quantum efficiency (QE), or the collection efficiency of the device, indicates how efficiently the optical signal is converted to electrical. As carriers are generated as photons are absorbed, only a certain percentage will be collected and sensed as photocurrent. Intuitively, QE is the number of sensed charge carriers per second per incident photon per second, or

$$\eta = \frac{J_p}{q\phi} \tag{1.2}$$

where η, J_p, J_d, q, and ϕ represent QE, photocurrent density[2], dark current density, electron charge, and incident photon flux. There are a few considerations when discussing photocurrent. First, photocurrent relies on photons being transmitted into the device; any reflection induced by surfaces that are not normally aligned to the incident radiation, or other unique geometrical features, may degrade QE. Second, the device must have a sufficient thickness in order to absorb a significant fraction of the radiation. If the device is too thin, light may be transmitted out of the device and not sensed as signal. Typically devices will have thickness that are at least two times

[2]Note, in this equation J_p is the photocurrent density due to an incident photon flux. When measuring the current under illumination, the total signal includes J_p and the dark current J_d. To obtain this value in practice it may be required to remove the dark current. In most cases, $J_p \gg J_d$, so J_d is negligible, but not always.

their characteristic absorption length, α^{-1} where α is the absorption coefficient of the material. Last, once carriers are generated they must be sensed as photocurrent before recombining. The inverse processes of Fig. 1·3(a), (c), and (d), or recombination mechanisms, can annihilate excess carriers as the material returns to equilibrium.

1.3.3 Specific detectivity

It is often required to compare the performance of devices that employ different materials and architectures. A useful metric that combines the dark current, or noise, of the detector with the quantum efficiency is specific detectivity, or D^*. D^* is defined as (Vincent et al., 2015)

$$D^* = \frac{\mathcal{R}\sqrt{A_d \Delta f}}{\mathcal{N}}, \tag{1.3}$$

where $\mathcal{R}$, $\mathcal{N}$, A_d, and Δf are the responsivity of the detector with units A/W, the noise, detector area, and system noise bandwidth. The responsivity is the amount of output current (or voltage) per incident optical power, and can be related back to the quantum efficiency and photocurrent by

$$\mathcal{R} = \frac{\eta \lambda q}{hc} = \frac{\eta \lambda}{1.24}, \tag{1.4}$$

with the latter assuming units of micrometers for the wavelength. Photovoltaic devices are often shot-noise limited (Rogalski, 2010, Vincent et al., 2015, Sze and Ng, 2006). Shot-noise is due to the random nature of carrier generation within the device, and is given by (Vincent et al., 2015, Rogalski, 2010)

$$i_N^2 = 2 I_d q \Delta f. \tag{1.5}$$

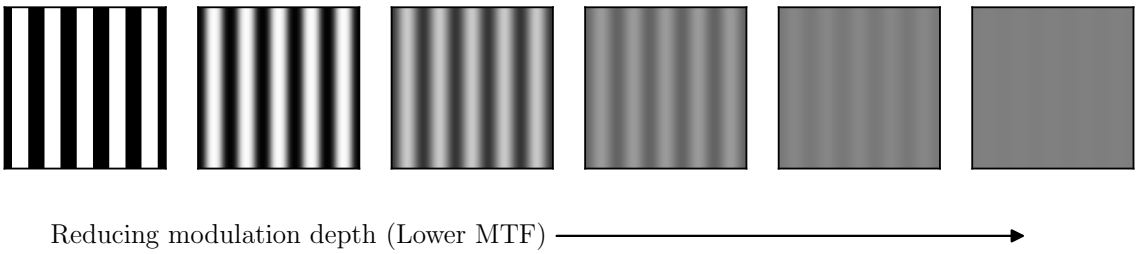

Figure 1·4: Bar target with increased aliasing, representing a sensor with decreasing MTF.

Assuming a system noise bandwidth of 1 Hz, we can write D^* as[3]

$$D^* = \frac{\eta\lambda}{1.24\sqrt{2qJ_d}}, \tag{1.6}$$

with units of $\mathrm{cm\,Hz^{1/2}\,W^{-1}}$ (Jones) when q, J_d, and λ are expressed in coulombs, $\mathrm{A\,cm^{-2}}$, and µm respectively. D^* is instrumental in assessing performance tradeoffs in a standardized way.

1.3.4 Modulation transfer function

The principal goal of any imaging system is to accurately reproduce a scene with high fidelity. The modulation transfer function (MTF) is a common parameter used to quantify the degree of aliasing that an imaging system introduces in the output. The MTF describes the system's response to sinusoids with varying spatial frequencies and is defined as the Fourier transform of the system's impulse response (Boreman, 2001). Shown in Fig. 1·4, the lower the MTF the higher degree of aliasing that is introduced in the target. The aliasing is a byproduct of the discrete sampling ability of an image sensor. Since the sensor is comprised of discrete imaging elements, there is a fundamental limit to the spatial frequencies that the sensor can detect. This is

[3]This expression will overestimate the actual specific detectivity; by assuming shot-noise dominant we are neglecting other noise sources, such as Johnson-Nyquist.

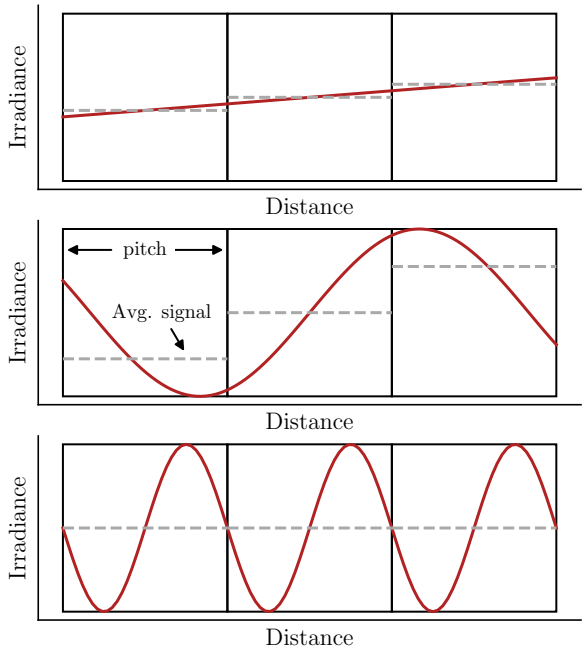

Figure 1·5: Illustration of a three pixel sensor subject to sinusoidal irradiance profiles of varying spatial frequency. Once the irradiance's spatial frequency exceeds the reciprocal of the pixel's pitch, all contrast is lost (Boreman, 2001).

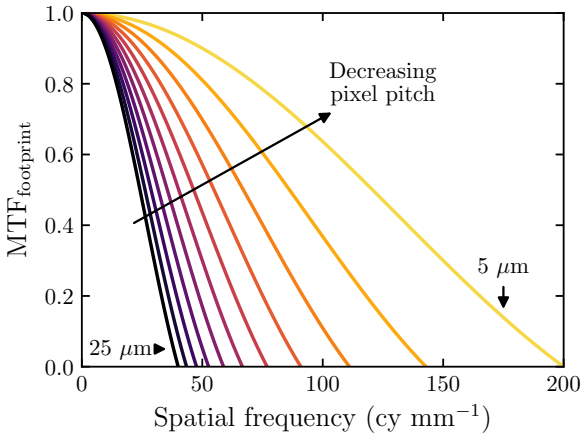

Figure 1·6: The ideal footprint of a image sensor with square pixels. As the pixel dimensions are reduced, the MTF improves (Boreman, 2001).

shown in Fig. 1·5 where the size of the pixels sets a limit on the spatial frequency content that a sensor can preserve. For low spatial frequencies, the pixels' average signal accurately represents the irradiance profile. As the spatial frequency increases, and approaches the reciprocal of the pixel's pitch, all contrast is lost in the image.

The MTF in an optical system is the product of each subcomponent's MTF, (Boreman, 2001)

$$\mathrm{MTF}_{\mathrm{system}} = \mathrm{MTF}_{\mathrm{optics}} \times \mathrm{MTF}_{\mathrm{sensor}} \times \mathrm{MTF}_{\mathrm{electronics}} \times \cdots$$

In this work we are concerned with the MTF of the image sensor, which can be written as a product the detector footprint multiplied by the sources of MTF degradation, (Pinkie et al., 2013)

$$\mathrm{MTF}_{\mathrm{sensor}} = \mathrm{MTF}_{\mathrm{footprint}} \times \mathrm{MTF}_{\mathrm{electrical\,crosstalk}} \times \mathrm{MTF}_{\mathrm{optical\,crosstalk}}.$$

The footprint of the detector represents the ideal response of the sensor, and is determined by the size and geometry of the imaging elements. The case of a rectangular array of pixels has the analytic solution (Boreman, 2001)

$$\mathrm{MTF}_{\mathrm{footprint}}(\xi, \eta) = \left| \frac{\sin(\pi \xi w_x)}{\pi \xi w_x} \right| \left| \frac{\sin(\pi \eta w_y)}{\pi \eta w_y} \right|, \tag{1.7}$$

where ξ and η are the spatial frequencies in the x and y directions considering a top-down view of an two-dimensional array of pixels, and w_x and w_y are the pixel dimensions. One method of improving MTF involves reducing the size of the pixels, as shown in Fig. 1·6. However, while the hypothetical limiting $\mathrm{MTF}_{\mathrm{footprint}}$ may improve with shrinking pixel size, there may be additional technological challenges to overcome to realize it. For example, smaller devices are inherently more complex to fabricate as lithography tolerances become tighter, or there may also be additional sources of MTF degradation from increasing inter-pixel crosstalk.

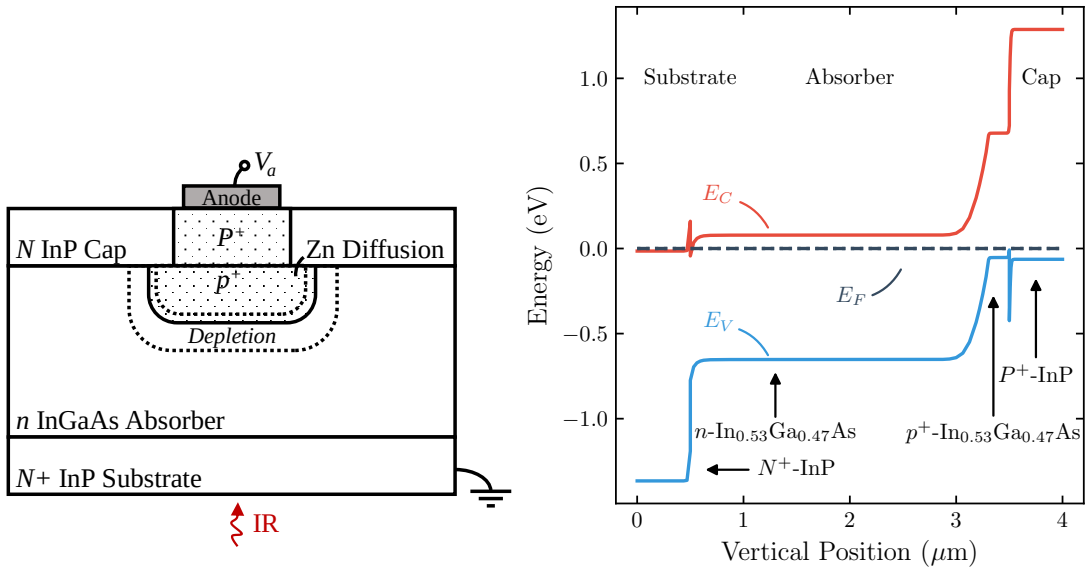

Figure 1·7: (Left) Exemplar photodiode using the double layer planar heterojunction architecture. (Right) Equilibrium band diagram vertically through the structure at 300K. The absorber and cap background doping densities are 10^{16} cm^{-3}, the p-type diffused region is assumed abrupt with constant doping density of 10^{18} cm^{-3}.

1.4 Photodetector architectures discussed in this dissertation

There are two types of pixel architectures considered in this work. The first is the double layer planar heterojunction (DLPH). The DLPH architecture has been successfully used with several infrared materials, including $In_{1-x}Ga_xAs$ and $Hg_{1-x}Cd_xTe$ (DeWames et al., 2015, Reine, 2001). A schematic of the architecture for a SWIR $In_{0.53}Ga_{0.47}As$ photodiode is shown along with the corresponding vertical band diagram in Fig. 1·7. The architecture is similar for MCT devices, but instead use separate compositions to form the heterojunctions. The following describes at a high-level the process for creating the photodiode. Beginning with a substrate, thin layers of semiconducting material are deposited. The first layer is often a buffer (not shown) intended to improve the structural and electrical properties of subsequent layers by

reducing dislocation propagation into the active devices layers in lattice-mismatched materials and lowering defect densities. Next the absorber is grown epitaxially. Ordinarily the absorber's thickness is chosen based on the characteristic absorption length. Finally, a large bandgap cap layer is grown. The cap increases the critical thickness of the absorber to allow thicker material to be grown and acts as a native passivation for the low bandgap absorber. After the epitaxial stack is grown, the diode is subsequently formed. For $In_{0.53}Ga_{0.47}As$, zinc is diffused through a diffusion mask to form P+-InP and p^+-$In_{0.53}Ga_{0.47}As$ usually by a gas source ampoule approach (Ettenberg et al., 1999). As noted by other authors, it is crucial that the p^+ diffusion reaches the $In_{0.53}Ga_{0.47}As$ absorber to maintain high quantum efficiency at low biases (DeWames et al., 2015). The final steps include depositing an insulating passivation layer and ohmic contact formation. For FPAs the photodiode arrays are hybridized to a separate ROIC, through an indium bump bond technique.

The DLPH pixel operates in the same way as a typical p-n photodiode. The photodiode is operated under small reverse bias, and the resulting change in current under illumination is measured. Most of the absorption occurs in the absorbing layer. Electron-hole pairs are generated; if they are generated within a diffusion length of the depletion region, they may diffuse to the depletion region where the electric field will separate them. The minority carriers—in this case holes—will drift through the depletion layer and are collected at the anode and sensed as photocurrent (Sze and Ng, 2006). A notable feature of the DLPH architecture is a large valence band barrier to holes at the substrate-absorber interface, shown in Fig. 1·7. This is an important boundary condition that leads to a higher QE when compared to a p-n homojunction device as holes are repelled from the interface towards the depletion region. These details are discussed at length in Chapter 3.

The second photodiode architecture discussed in this dissertation is the barrier de-

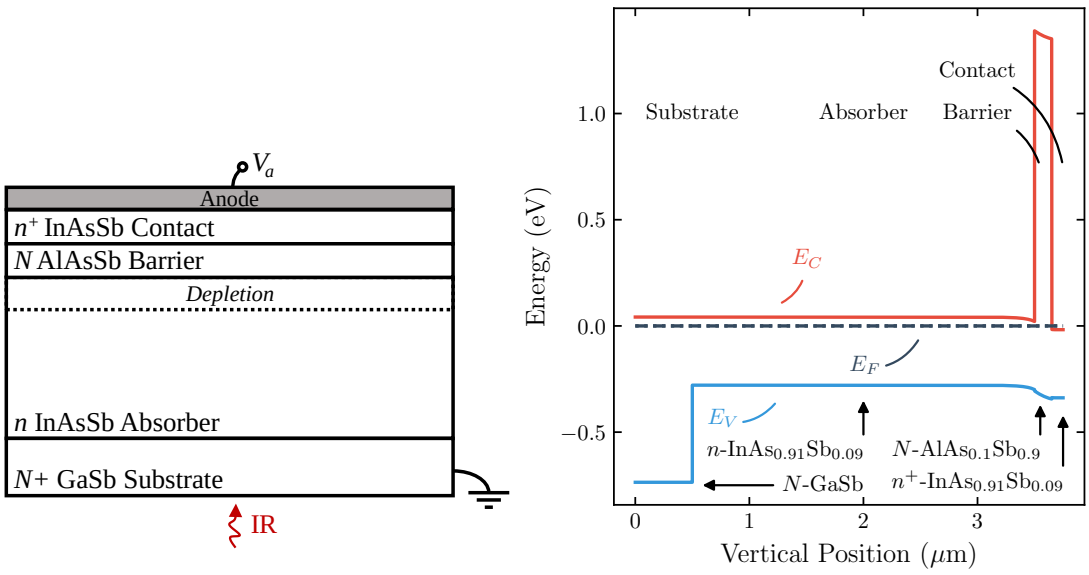

Figure 1·8: (Left) Exemplar epitaxial stack for an InAsSb-based barrier photodiode with n-type contact and N-type barrier layers. (Right) Equilibrium band diagram vertically through the structure. Note, at equilibrium the absorber-barrier interface is accumulated, a small reverse bias is required to slightly deplete the interface. The absorber, barrier, and contact doping densities are 10^{15}, 10^{15}, and 10^{17} cm^{-3} respectively. The substrate is idealized GaSb without a conduction band offset between the substrate and absorber.

tector. The original proposal for a barrier-style detector is often attributed to White, where a large bandgap material is sandwiched between two small gap semiconducting layers with zero valence band offset between the layers (White, 1987). This effectively creates a unipolar unity gain detector as the large barrier in the conduction band blocks majority carrier flow, preventing any photoconductive gain due to the asymmetric transit times of the photogenerated electron and holes. In 2006 Maimon and Wicks demonstrated a MWIR detector, which they labeled "nBn", using InAs absorbing and contact layers, and an AlAsSb barrier (Maimon and Wicks, 2006). Since then, barrier-style devices have enjoyed increasing popularity and success, especially for MWIR and LWIR, due to the simple fabrication process and several advantages over other III-V junction-based architectures. The most notable improvements include suppression of GR currents from the absorber and *in situ* passivation of the absorber surface. Another attractive feature of the barrier architecture is the flexibility in design choices. The design readily lends itself for dual-band imaging by replacing the contact layer with a smaller bandgap material; then, longer wavelengths will pass through the absorber and be absorbed in the contact layer. By reversing the polarity of the applied voltage we can target the separate spectral band. Moreover, the doping density in the barrier and contacts can be replaced with p-type. In this case, the absorber-barrier interface will start in slight depletion, reducing the required operating voltage. The n-type case has the advantage of optimizing the operating voltage around minimizing the dark current since the absorber-barrier interface starts accumulated (Klipstein, 2008).

In this work, we focus on a barrier-style device using $InAs_{1-x}Sb_x$ as the absorbing and contact material. A typical epitaxial stack is shown in Fig. 1·8. Growth will begin with a buffer layer (not shown) on a GaSb substrate. Afterwards, the absorber layer (AL), barrier layer (BL), and contact layers (CL) are grown. The final steps for

the device fabrication include a mesa delineation etch to isolate the individual pixels, passivation of the mesa sidewalls, ohmic contact formation, and ROIC hybridization. Here, we focus on $InAs_{1-x}Sb_x$ with 9% Sb content, but there has also been progress in extending the cutoff wavelengths deeper into the MWIR and LWIR with higher Sb compositions (Lin et al., 2015).

The operation of a barrier photodetector is very similar to a typical photodiode. Absorption occurs in the absorber and the optically generated carriers diffuse towards the absorber-barrier interface. As shown in the band diagram in Fig. 1·8, for low to moderate doping densities the barrier layer will be fully depleted, and the minority carriers are collected by the field and sensed as additional photocurrent. The barrier device is often referred to as a unipolar photodetector of unity gain, due to the presence of a large barrier in the conduction band preventing photoconductive gain due to the asymmetric transit times of the optically generated carriers (Sze and Ng, 2006, Klipstein, 2008). The detector is a minority carrier device and is analogous to back-to-back heterojunction diodes (Reine et al., 2014).

1.5 Scope and organization of this research

Since their first appearance in the late 1980's and early 1990's infrared focal plane arrays have seen a consistent drive for reducing their size, weight, power consumption, and cost. One approach that can achieve all of these requirements is reducing the pixel size. Indeed, shown in Fig. 1·9 pixel sizes have steadily decreased over time, with recent demonstrations of 5 µm pixel pitch FPAs for SWIR and 10 to 15 µm pitches for MWIR and LWIR[4] (MacDougal et al., 2020, Yuan et al., 2012, Rogalski et al., 2016). One advantage, mentioned previously, is that smaller pixels are able to capture more closely spaced object with higher quality as the detector's $MTF_{footprint}$ would improve.

[4]This is mostly due to dual-band detectors being fabricated at MWIR pitches (Rogalski et al., 2016).

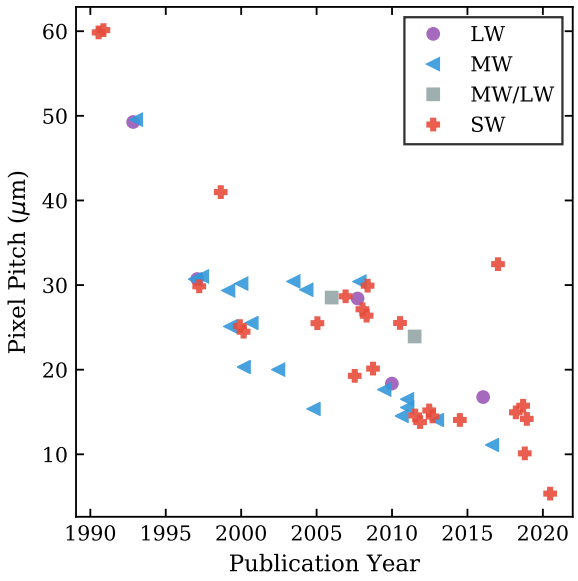

Figure 1·9: Pixel size over the past 30 years from various producers of infrared imaging systems. A small amount of jitter has been added to the points to improve the clarity of overlapping data.

Assuming $f/\# = 1.4$ for SWIR ($\lambda = 1$ µm) and for MWIR ($\lambda = 4$ µm), the image sensor would require pixel pitches smaller than 3.4 µm and 13.6 µm for the system's MTF to transition from detector-limited to optics-limited (Holst, 2007, Rogalski et al., 2016).

There are additional advantages for moving to smaller pixels aside from improving image quality. One is the ability to leverage an economy of scale, where as pixel pitches are reduced, focal plane arrays with the same resolution have less material footprint, allowing for more focal plane arrays per wafer. Conversely, higher resolutions are achievable with similar die sizes. Furthermore, if a smaller die size is available, the size of the optical elements can be reduced, helping in reducing the total size and weight of a portable system (Rogalski et al., 2016).

To this end, this dissertation offers several contributions to the development of infrared photodetectors. In the first half, we review challenges at the detector-level when moving to smaller pixels. Using a three-dimensional numerical model for simu-

lating the performance of $In_{0.53}Ga_{0.47}As$ DLPH photodiodes, we discuss the impact on pixel size reduction in terms of dark current, D^*, and MTF, and introduce a strategy intended to mitigate any adverse effects. Though the results and discussion are based on $In_{0.53}Ga_{0.47}As$, most of the trends, methodology, and analysis can be applied to other material systems and devices.

The second half of the dissertation continues to build on existing simulation techniques for infrared photodetectors. Using a combination of analytical and numerical modeling we offer a comprehensive overview of the capacitance-voltage (C-V) characteristics of barrier detectors. The results of an extensive parametric study of nBn devices with $N-$ and $P-$type barrier layers are explained by the underlying device physics, and shown to agree well with metal-oxide-semiconductor capacitor theory for certain parameter choices. From this study, it was determined that the C-V profile of barrier detectors can offer a deeper insight into the underlying photodiode architecture than what is typically expected from C-V profiling.

Finally, we present several applications of applying artificial neural networks (ANNs) to infrared detector characteristics. First, using the insight from the C-V modeling we demonstrate that ANNs are able to be used as a surrogate model for the C-V characteristics of nBn devices, and are useful to visualize and understand the role of each feature in multi-dimensional parameter spaces. Second, we demonstrate that convolutional neural networks (CNNs) can be trained to predict the underlying device architectural properties when given a C-V profile as input. Then, the CNN can be used as an enhanced analysis tool for C-V characteristics to gain more information about the device than what would be possible through conventional C-V profiling. Finally, using ANNs we explore the tradeoffs between QE and dark current for competing device designs, and assess the tolerance of a given design to uncertainties in the design parameters.

Chapter 2

Numerical Simulation Methods

This chapter provides a brief overview of the methods used to generate the results in this dissertation. First, we discuss the finite-difference time-domain (FDTD) method for simulating the electromagnetic field throughout electro-optical devices. Second, we introduce the set of equations that comprise the drift-diffusion method for simulating carrier transport in semiconductor devices. Finally, we provide an overview of the simulation strategy to compute the figures of merit.

2.1 Finite-difference time-domain

2.1.1 First-order finite differences in one dimension

Consider the Taylor series expansion of a function $f(x)$ about x_i given by

$$f(x) = f(x_i) + \frac{f'(x_i)}{1!}(x - x_i) + \frac{f''(x_i)}{2!}(x - x_i)^2 + \cdots + \frac{f^{(n)}(x_i)}{n!}(x - x_i)^n. \quad (2.1)$$

Evaluating (2.1) at $x = x_i + \Delta x$ gives

$$f(x_i + \Delta x) = f(x_i) + f'(x_i)\Delta x + \frac{f''(x_i)}{2}\Delta x^2 + \cdots + \frac{f^{(n)}(x_i)}{n!}\Delta x^n. \quad (2.2)$$

Rearranging (2.2) to solve for $f'(x_i)$ results in the well-known first-order right hand side finite-difference formula

$$f'(x_i) \approx \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x} + \mathcal{O}(\Delta x)$$

where $\mathcal{O}(\Delta x)$ is the truncation error introduced by dropping the higher order terms in the approximation. Following similar logic, evaluating (2.1) at $x = x_i - \Delta x$ gives

$$f(x_i - \Delta x) = f(x_i) - f'(x_i)\Delta x + \frac{f''(x_i)}{2}\Delta x^2 + \cdots + \frac{f^{(n)}(x_i)}{n!}(-\Delta x)^n, \tag{2.3}$$

and likewise solving for $f'(x_i)$ yields the left-sided finite-difference formula

$$f'(x_i) \approx \frac{f(x_i) - f(x_i - \Delta x)}{\Delta x} + \mathcal{O}(\Delta x).$$

By subtracting (2.2) and (2.3), and again solving for $f'(x_i)$, we derive the central difference formula

$$f'(x_i) \approx \frac{f(x_i + \Delta x) - f(x_i - \Delta x)}{2\Delta x} + \mathcal{O}(\Delta x^2). \tag{2.4}$$

Notice that the truncation error is now of the order $\mathcal{O}(\Delta x^2)$ due to the odd terms canceling out. The central difference equation enables an approximate solution of a function's derivative at a point by simply using the functions values at the neighboring grid points. The central difference also reduces the truncation error quadratically with reduction in Δx. We will see later that this is useful for solving initial and boundary condition problems containing partial differential equations.

2.1.2 Maxwell's equations

Using the equations from the previous section, and following (Taflove and Hagness, 2005), we derive the finite-difference time-domain equations for finding the wave solution to Maxwell's equations. Maxwell's equations are ubiquitous in science and engineering for numerous applications requiring understanding of the electromagnetic properties and performance of electronic circuits, motors, antennas, and of course,

sensors. In differential form, Maxwell's equations are written as

$$\frac{\partial \boldsymbol{B}}{\partial t} = -\boldsymbol{\nabla} \times \boldsymbol{E} - \boldsymbol{M}, \tag{2.5}$$

$$\frac{\partial \boldsymbol{D}}{\partial t} = \boldsymbol{\nabla} \times \boldsymbol{H} - \boldsymbol{J}, \tag{2.6}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{D} = 0, \tag{2.7}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0, \tag{2.8}$$

where (2.5) is Faraday's law, (2.6) is Ampere's law, and (2.7) and (2.8) are Gauss' laws for electric and magnetic fields respectively. The symbols appearing in (2.5)–(2.8) are defined as:

$\boldsymbol{B}$	-	magnetic flux density (Wb m^{-2})
$\boldsymbol{E}$	-	electric field (V / m)
$\boldsymbol{M}$	-	magnetic current density (V m^{-2})
$\boldsymbol{D}$	-	electric flux density (C m^{-2})
$\boldsymbol{H}$	-	magnetic field (A m^{-1})
$\boldsymbol{J}$	-	electric current density (A m^{-2})

$\boldsymbol{D}$ and $\boldsymbol{H}$ are also commonly referred to as the electric displacement field and magnetizing field respectively, and in materials with frequency, direction, and field independent electronic and magnetic properties can be given by

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} = \varepsilon_0 \varepsilon_r \boldsymbol{E} \tag{2.9}$$

$$\boldsymbol{H} = \mu \boldsymbol{B} = \mu_0 \mu_r \boldsymbol{B} \tag{2.10}$$

where ε_0, ε_r, μ_0, and μ_r are the free space permittivity, relative permittivity, free space magnetic permeability, and the relative magnetic permeability. $\boldsymbol{J}$ and $\boldsymbol{M}$ represent the electric and magnetic current densities, and are proportional to the electric and magnetic fields in addition to any sources of electromagnetic flux, or,

$$\boldsymbol{J} = \boldsymbol{J}_{\text{source}} + \sigma \boldsymbol{E} \tag{2.11}$$

$$\boldsymbol{M} = \boldsymbol{M}_{\text{source}} + \sigma^* \boldsymbol{H}. \tag{2.12}$$

where σ and σ^* are the electric conductivity and magnetic loss. Substituting (2.9)–(2.12) into (2.5) and (2.6) gives Maxwell's curl equations:

$$\frac{\partial \boldsymbol{H}}{\partial t} = -\frac{1}{\mu}\boldsymbol{\nabla} \times \boldsymbol{E} - \frac{1}{\mu}(\boldsymbol{M}_{\text{source}} + \sigma^* \boldsymbol{H}) \tag{2.13}$$

$$\frac{\partial \boldsymbol{E}}{\partial t} = \frac{1}{\varepsilon}\boldsymbol{\nabla} \times \boldsymbol{H} - \frac{1}{\varepsilon}(\boldsymbol{J}_{\text{source}} - \sigma \boldsymbol{E}). \tag{2.14}$$

These equation comprise a set of six coupled scalar equations that when solved, determine the propagation of electromagnetic radiation through a medium. Expanding the vector notation in Cartesian coordinates gives

$$\frac{\partial H_x}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y} - M_{x,\text{source}} - \sigma^* H_x\right) \tag{2.15}$$

$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} - M_{y,\text{source}} - \sigma^* H_y\right) \tag{2.16}$$

$$\frac{\partial H_z}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x} - M_{z,\text{source}} - \sigma^* H_z\right) \tag{2.17}$$

$$\frac{\partial E_x}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - J_{x,\text{source}} - \sigma E_x\right) \tag{2.18}$$

$$\frac{\partial E_y}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - J_{y,\text{source}} - \sigma E_y\right) \tag{2.19}$$

$$\frac{\partial E_z}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - J_{z,\text{source}} - \sigma E_z\right). \tag{2.20}$$

2.1.3 Yee Algorithm

Equations (2.15)–(2.20) comprise the system of first order partial differential equations that must be solved to obtain a solution for a propagating electromagnetic wave. In 1966, Kane Yee published his approach to this problem, which has since

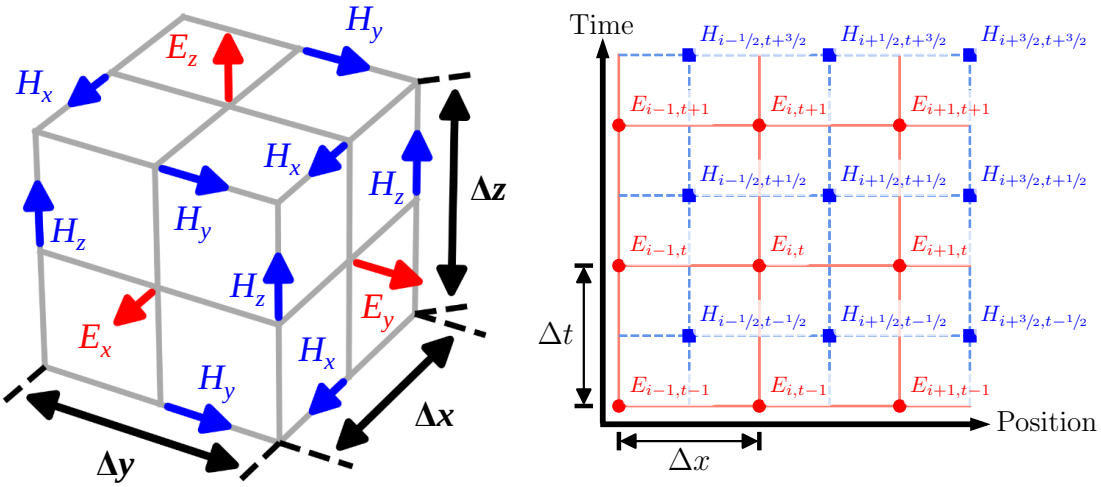

Figure 2·1: (Left) The arrangement of electric and magnetic field components on a cubic grid. (Right) Time and space discretization of the Yee algorithm. After (Taflove and Hagness, 2005).

received wide adoption as an accurate, numerically stable, and robust solution (Yee, 1966). This section covers the basic principles of the approach; for more information regarding either FDTD, or the Yee algorithm, readers are encouraged to look elsewhere (Taflove and Hagness, 2005).

Consider a discretized space three dimensions where each cell has a constant width in each direction, Δx, Δy, and Δz in the x, y, and z directions respectively. In the Yee algorithm the electric and magnetic field components are interleaved throughout the unit cell as shown in Fig. 2·1. Similarly, (2.15)–(2.20) have both spatial and temporal components, as such we assume a constant time step of Δt. A similar interleaving approach is used to distribute the electric and magnetic fields to implement a leapfrog algorithm in time, also shown in Fig. 2·1. For consistency, we adopt the same shorthand notation in (Taflove and Hagness, 2005) for each scalar component of $\boldsymbol{H}$ and $\boldsymbol{E}$. For example, the value of E_x at $t = t_n$ and position $(x, y, z) = (x_i, y_j, z_k)$ is written $E^n_{x,i,j,k}$. Values of E_x at time $t = t_n + \Delta t$ and po-

sition $(x, y, z) = (x_i + \Delta x, y_j + \Delta y, z_k + \Delta z)$ is written $E^{n+1}_{x,i+1,j+1,k+1}$. To illustrate the approach, we start by applying the central difference formula (2.3) to (2.18) and obtain[1]

$$\frac{E^{n+1/2}_{x,i,j,k} - E^{n-1/2}_{x,i,j,k}}{\Delta t} = \frac{1}{\varepsilon}\left[\frac{H^n_{z,i,j+1/2,k} - H^n_{z,i,j-1/2,k}}{\Delta y} - \frac{H^n_{y,i,j,k+1/2} - H^n_{y,i,j,k-1/2}}{\Delta z} - J^n_{x,i,j,k} - \sigma E^n_{x,i,j,k}\right]. \quad (2.21)$$

The term on the left-hand side includes the value of E_x at the next time step, $E^{n+1/2}_{x,i,j,k}$, and the previous value, $E^{n-1/2}_{x,i,j,k}$. All the terms on the right-hand side are at the current time, t_n. Only the previous value of $E_{x,i,j,k}$ at $t_{n-1/2}$ has been computed, so the value of $E^n_{x,i,j,k}$ is assumed to be the average of $E^{n+1/2}_{x,i,j,k}$ and $E^{n-1/2}_{x,i,j,k}$. Making this substitution and rearranging (2.21) yields

$$E^{n+1/2}_{x,i,j,k} = \left(\frac{1 - \sigma\Delta t/2\varepsilon}{1 + \sigma\Delta t/2\varepsilon}\right) E^{n-1/2}_{x,i,j,k} + \left(\frac{\Delta t/\varepsilon}{1 + \sigma\Delta t/2\varepsilon}\right) \cdot \left(\frac{H^n_{z,i,j+1/2,k} - H^n_{z,i,j-1/2,k}}{\Delta y} - \frac{H^n_{y,i,j,k+1/2} - H^n_{y,i,j,k-1/2}}{\Delta z} - J^n_{x,i,j,k}\right). \quad (2.22)$$

Equations similar to (2.22) are used to update the values of the components of **E** and **H** throughout the simulation domain as time is stepped during the solution. The elegance of Yee's solution to Maxwell's equations cannot be understated; the FDTD approach is favorable over solutions of scalar wave equations as it includes both electric and magnetic field components, and it can be shown that the Yee unit cell satisfies Gauss' laws (2.7) and (2.8) (Taflove and Hagness, 2005).

[1]Notice that the central differences for both the partial derivative with respect to time and space centered around $\pm 1/2\Delta t$, $\pm 1/2\Delta y$, and $\pm 1/2\Delta z$, instead of $\pm\Delta x$ in Eq. (2.4).

2.2 Drift-diffusion model

In this section, we review the equations that are involved when using the drift-diffusion (DD) method to simulate carrier transport in semiconductor devices.

2.2.1 Continuity equations

Consider a thin slice of semiconducting material located at x with thickness dx. The rate of change of minority carriers in this slice is then equal to the difference between the inward and outward flux of carriers, plus the difference of the number of created and annihilated carriers, or,

$$\frac{\partial p}{\partial t}Adx = \frac{J_p(x) - J_p(x+dx)}{q}Adx + (G_p - R_p)Adx, \tag{2.23}$$

where p, J_p, G_p, and R_p are the hole concentration, hole current density, hole generation rate, and hole recombination rate. Expanding $J_p(x+dx)$ using a Taylor's series, (2.23) becomes

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\frac{\partial J_p}{\partial x} + G_p - R_p. \tag{2.24}$$

Likewise for electrons,

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} + G_n - R_n. \tag{2.25}$$

The current densities for the carriers are comprised of two components. The first is due to an external force enacting on the carriers. When a voltage is applied across a semiconductor, an electric field is created that provides energy to the system. Electrons and holes gain a drift velocity based on the direction of the external field, resulting in a drift current given by

$$J_{n,\text{drift}} = nq\mu_n\mathcal{E}, \tag{2.26}$$

and

$$J_{p,\text{drift}} = pq\mu_p\mathcal{E}, \tag{2.27}$$

where μ_n, μ_p, and $\mathcal{E}$ are the electron and hole mobilities and the electric field. The second component of current is due to the random motion of free carriers. To first order approximation, the net current as carriers randomly move through the crystal lattice is

$$J_{p,\text{diffusion}} = -qD_p\frac{\partial p}{\partial x}, \tag{2.28}$$

and

$$J_{n,\text{diffusion}} = qD_n\frac{\partial n}{\partial x}, \tag{2.29}$$

for holes and electrons respectively. The diffusivities, D_n and D_p, are given by the Einstein relations

$$D_n = \frac{kT\mu_n}{q},$$
$$D_p = \frac{kT\mu_p}{q}.$$

These two current components are the diffusion currents, and indicate that gradient of the carrier concentrations follows the direction of high to low concentration. Substituting the drift and diffusion current equations into (2.24) and (2.25) yields

$$\frac{\partial p}{\partial t} = -\mu_p\mathcal{E}\frac{\partial p}{\partial x} - \mu_p p\frac{\partial \mathcal{E}}{\partial x} + D_p\frac{\partial^2 p}{\partial x^2} + G_p - R_p, \tag{2.30}$$

$$\frac{\partial n}{\partial t} = \mu_n\mathcal{E}\frac{\partial n}{\partial x} + \mu_n n\frac{\partial \mathcal{E}}{\partial x} + D_n\frac{\partial^2 n}{\partial x^2} + G_n - R_n. \tag{2.31}$$

or in three dimensions

$$\frac{\partial p}{\partial t} = \nabla\cdot(-\mu_p p\mathcal{E} + qD_p\nabla p) + G_p - R_p, \tag{2.32}$$

$$\frac{\partial n}{\partial t} = \nabla \cdot (\mu_n n \mathcal{E} + q D_n \nabla n) + G_n - R_n. \tag{2.33}$$

Equations (2.32) and (2.33) when combined with Poisson's equation,

$$\nabla \cdot \nabla \phi = -\frac{q}{\varepsilon}(n - p + N_A - N_D), \tag{2.34}$$

comprise the set of coupled equations that are used in the drift-diffusion method. For performing the numerical calculations in this work we use the drift-diffusion simulator provided by Synopsys Sentaurus TCAD. The software uses a direct numerical simulation method based on finite-volumes to solve (2.32)–(2.34) for the unknowns, n, p, and ϕ.

2.3 Simulation strategy

As mentioned previously, the Sentaurus TCAD software suite is used to perform both the finite-difference time-domain and drift-diffusion simulations. This section serves as concise overview of the simulation approach adopted in this work. More details on the development of the approach, methodology, and validation against various infrared photodetectors are offered in other works (Bellotti and D'Orsogna, 2006, Pinkie et al., 2013, Wichman et al., 2014).

2.3.1 Quantum efficiency

The simulation workflow begins with discretizing the device of interest. Special care must be taken to ensure that the grids are dense enough in important regions, such as near contacts, junctions, interfaces, and regions with varying quantities such as electric fields, composition, or doping. For FDTD simulations, a tensorial mesh is used with cell sizes that depend on the material and wavelength of the simulation as

$$\text{cellsize} = \frac{\lambda \cdot (\text{nodes per wavelength})}{\sqrt{n^2 + k^2}},$$

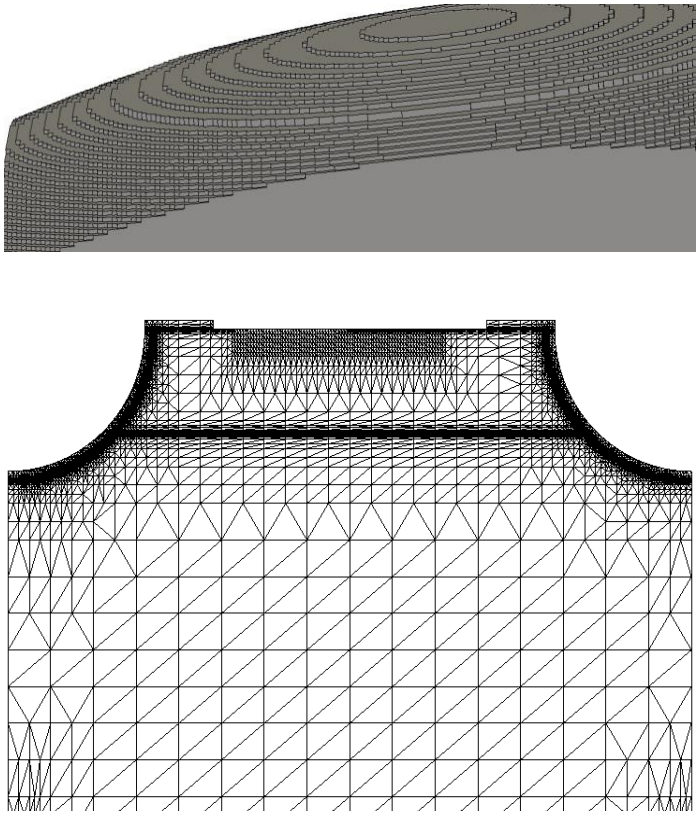

Figure 2·2: (Top) Tensorial grid for FDTD simulation of a device with a monolithically integrated microlense. (Bottom) Finite-element tetrahedral grid for drift-diffusion simulation of a mesa-style photodiode.

where n and k are the index of refraction and extinction coefficient. The number of nodes per wavelength generally must be higher than 10, but can be increased as needed to improve stability or accuracy. In this work the nodes per wavelength is between 15 and 25 for three-dimensional calculations, and 50 for one- or two-dimensional simulations. For drift-diffusion simulations a finite-element mesh composed of tetrahedral elements is used. Examples of these meshes are shown in Fig. 2·2.

First, a FDTD simulation is used to calculate the electromagnetic field throughout the device. When calculating the QE, we assume uniform plane wave illumination. The optical generation rate is found from the intensity distribution through the device

$$G_{opt} = \frac{-\nabla \cdot \boldsymbol{S_{av}}}{E_{ph}} = \frac{\sigma |\boldsymbol{E}|^2}{2E_{ph}},$$

where E_{ph} is the photon energy and $\boldsymbol{S_{av}}$ is the time-average Poynting vector (Synopsys, 2017b). At the conclusion of the simulation the optical generation is interpolated onto a finite-element mesh for subsequent drift-diffusion simulation. Two separate drift-diffusion simulations are performed. The first is under dark conditions to measure the dark current. The second uses a previously computed optical generation profile to calculate the photo current for a given wavelength and photon flux. From these quantities the QE can be calculated.

2.3.2 Modulation transfer function

Simulating the MTF of a focal plane array has similarities with the measurement of the point spread function (PSF) of an image sensor, but on a smaller scale (Boreman, 2001). The basic premise to simulating the MTF is to measure the response of a pixel as a function of illumination source position; these mappings are referred to as spot-scan (SS) profiles. The spot-scan profile is built from measuring the photocurrent in the central pixel of a 3x3 array as a function of the position of a Gaussian beam

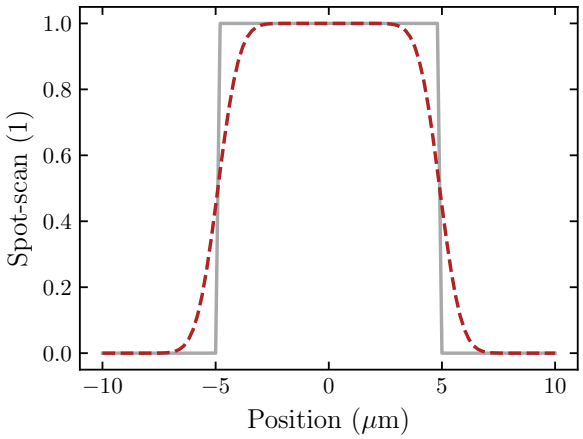

Figure 2·3: Idealized impulse response of a pixel with a Gaussian beam excitation. Solid line is the box function, dashed line is the convolution of the box function with a Gaussian pulse.

excitation. The total simulated MTF from the subsequent drift-diffusion simulations is then defined as the Fourier transform of this spot-scan profile

$$\mathrm{MTF_{DD}} = \mathcal{F}(\mathrm{SS_{photocurrent}(x, y)}). \tag{2.35}$$

Consider sweeping a beam across the center of a pixel array centered at $x = 0$ in one dimension from $x =$ -pitch to $x =$ +pitch. The ideal impulse response of the central pixel considering a point source is given by a box function of width equal to the pitch (Boreman, 2001). This intuitively describes the ideal behavior of a rectangular array; as the source is scanned off of the central pixel the signal should immediately drop to zero as all the signal should then be collected by the illuminated pixel. However, the simulated source has a spatial distribution whose normalized intensity at a given depth under normal incidence is given by

$$I_{\mathrm{GB}}(r) = e^{-2r^2/r_0^2}.$$

Shown in Fig. 2·3 is the convolution of the box function with the Gaussian intensity profile, which represents the ideal simulated response to a Gaussian excitation. Given

this information, the total simulated MTF is given by

$$\mathrm{MTF_{DD}} = \mathrm{MTF_{footprint}} \times \mathrm{MTF_{electrical\,crosstalk}} \times \mathrm{MTF_{optical\,crosstalk}} \times \mathrm{MTF_{GB}} \quad (2.36)$$

where $\mathrm{MTF_{electrical\,crosstalk}}$ is due to carriers diffusing from the pixel they were generated in to adjacent ones, $\mathrm{MTF_{optical\,crosstalk}}$ is due to photons scattering into neighboring pixels from any surface features such as mesa sidewalls or contacts, and $\mathrm{MTF_{GB}}$ is the contribution from the Gaussian beam excitation given by

$$\mathrm{MTF_{GB}} = \mathcal{F}(I_{\mathrm{GB}}).$$

Then, the true MTF of the detector can be recovered by

$$\mathrm{MTF_{sensor}} = \frac{\mathrm{MTF_{DD}}}{\mathrm{MTF_{GB}}} = \frac{\mathcal{F}(SS_{\mathrm{photocurrent}})}{\mathcal{F}(I_{\mathrm{GB}})}. \quad (2.37)$$

This correction for the spatial content of the Gaussian beam has been verified to yield accurate results (Pinkie et al., 2013).

We can also determine the contribution to the MTF from optical crosstalk from the simulations. During the FDTD simulations the optical generation throughout the device is calculated for each beam position. The total optical generation in the central pixel while the beam is centered at (x, y) is the volume integral of the optical generation rate,

$$G_{\mathrm{total}}(x, y) = \iiint G_{\mathrm{optical}} dV.$$

The Fourier transform of this set of total optical generation versus position gives the FDTD simulated MTF:

$$\mathrm{MTF_{FDTD}} = \mathcal{F}(SS_{\mathrm{generation}}). \quad (2.38)$$

Similar to the MTF from the photocurrent spot-scan, the MTF from the FDTD simulations has the Gaussian beam spatial content embedded in it, and must be corrected

$$\mathrm{MTF}_{\mathrm{FDTD}} = \mathrm{MTF}_{\mathrm{optical\,crosstalk}} \times \mathrm{MTF}_{\mathrm{GB}} \times \mathrm{MTF}_{\mathrm{footprint}}. \tag{2.39}$$

Hence, the MTF contributions from optical and electrical crosstalk can be obtained directly from the simulations by using (2.36):

$$\mathrm{MTF}_{\mathrm{electrical\,crosstalk}} = \frac{\mathrm{MTF}_{\mathrm{DD}}}{\mathrm{MTF}_{\mathrm{FDTD}}}, \tag{2.40}$$

$$\mathrm{MTF}_{\mathrm{optical\,crosstalk}} = \frac{\mathrm{MTF}_{\mathrm{FDTD}}}{\mathrm{MTF}_{\mathrm{GB}} \times \mathrm{MTF}_{\mathrm{footprint}}}. \tag{2.41}$$

Using this approach provides a way to identify which component limits the sensor's MTF.

Chapter 3

Diffusion current and ideal performance of junction-based photodiodes

This chapter reviews the solutions of the continuity equations with the goal of deriving analytic expressions for the dark current, photocurrent, and quantum efficiency of junction-based infrared photodetectors. In this chapter we are interested in the diffusion-limited performance, and focus on solutions to the continuity equations applied to the quasi-neutral regions of the photodiode. Following a similar procedure to Trezza et al., we also review the various boundary conditions that are present when considering arrays of photodiodes, and study how they affect these figures of merit (Trezza et al., 2011).

3.1 *p-n* junctions

The foundation of a basic photodiode is a p-n junction. Consider the structure shown in Fig. 3·1. In developing the following analytical models we assume that the junction is asymmetric: the p^+ region has a higher doping level than the n region, and is much

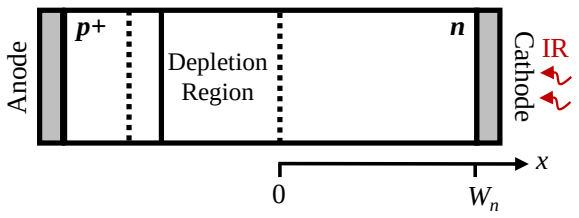

Figure 3·1: One-dimensional representation of a p-n junction.

thinner such that the p-n junction diffusion current is limited by the diffusion of holes from the n side. In one dimension, (2.32) when applied to the quasi-neutral n region becomes

$$\frac{\partial p(x)}{\partial t} = D_p \frac{\partial^2 p(x)}{\partial x^2} + G_p - R_p \tag{3.1}$$

where $p = p_{n0} + \Delta p$, p_{n0} is the equilibrium minority carrier concentration, and Δp is the excess carrier concentration caused by non-equilibrium conditions, such as an applied voltage or incident photon flux. The external generation rate, G_p, is zero under dark conditions. The recombination rate, R_p, under assumption of low injection, is

$$R_p = \frac{\Delta p}{\tau_p}, \tag{3.2}$$

where τ_p is the minority carrier lifetime given by the Matthiessen rule

$$\frac{1}{\tau_p} = \sum_{i}^{\text{Mechanisms}} \frac{1}{\tau_i} = \frac{1}{\tau_{\text{A}}} + \frac{1}{\tau_{\text{R}}} + \frac{1}{\tau_{\text{SRH}}} + ... \tag{3.3}$$

In this section we assume the Shockley-Read-Hall lifetime, τ_{SRH}, is long enough such that Auger and radiative recombination are dominant. The radiative and Auger lifetimes are given by

$$\tau_{\text{R}} = \frac{1}{BN_d}, \tag{3.4}$$

$$\tau_{\text{A}} = \frac{1}{CN_d^2}, \tag{3.5}$$

where B and C are the radiative and Auger coefficients. This discussion is based on using lattice-matched $In_{0.53}Ga_{0.47}As$ as the semiconductor material. A summary of the material models is provided in the appendix.

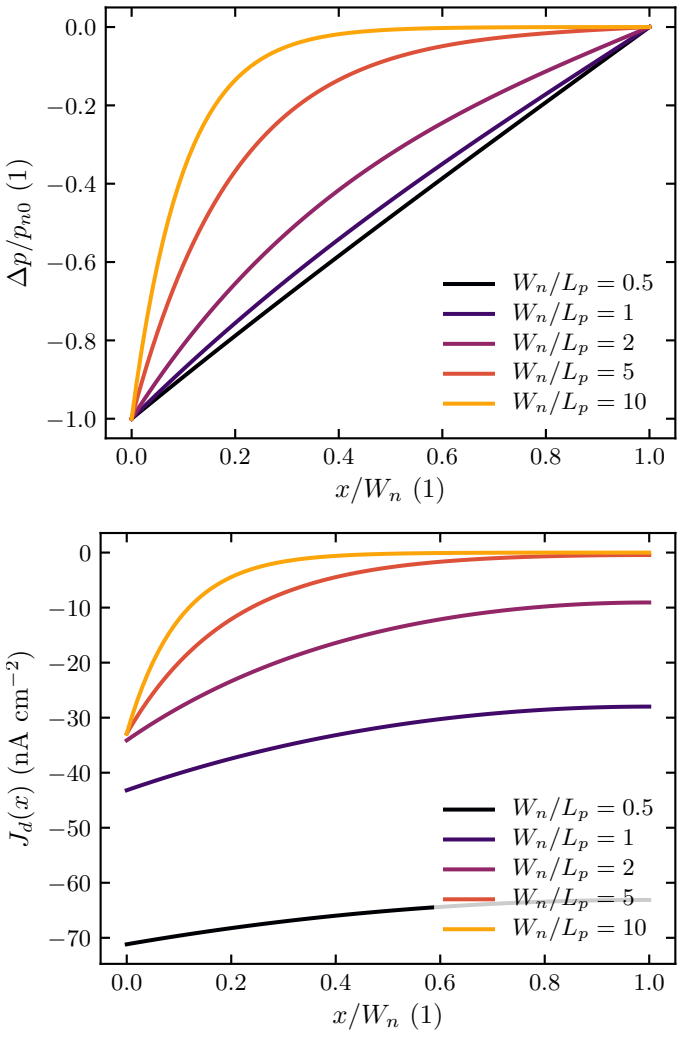

Figure 3·2: (Top) Excess minority carrier concentration and (bottom) current density under dark conditions for varied ratios of diffusion length and quasi-neutral region width.

3.1.1 Dark current density

By substituting $p(x) = p_{n0} + \Delta p(x)$ into (3.1) and assuming steady-state we obtain the following second-order homogenous differential equation for holes

$$D_p \frac{\partial^2 \Delta p}{\partial x^2} - \frac{\Delta p(x)}{\tau_p} = 0. \tag{3.6}$$

Following Fig. 3·1, the appropriate boundary conditions are (Muller and Kamins, 2003)

$$\Delta p(0) = p_{n0}\left(e^{qV/kT} - 1\right) \approx -p_{n0}, \tag{3.7}$$

$$\Delta p(W_n) = 0. \tag{3.8}$$

The first assumes a sufficiently large reverse bias, $|V_a| > kT/q$, such that the value of the total minority carrier concentration is zero at the edge of the depletion region. The second represents an ohmic contact that maintains the minority carrier concentration at its equilibrium value. The solution of (3.6) is

$$\Delta p(x) = -p_{n0} \operatorname{csch}\left(\frac{W_n}{L_p}\right) \sinh\left(\frac{W_n - x}{L_p}\right), \tag{3.9}$$

where L_p is the minority carrier diffusion length given by $\sqrt{D_p \tau_p}$. Recall that the current density is proportional to the gradient of the minority carrier concentration, or in this case,

$$\begin{aligned} J_p(x) &= -qD_p \frac{\partial \Delta p(x)}{\partial x} \\ &= -\frac{qp_{n0}L_p}{\tau_p} \operatorname{csch}\left(\frac{W_n}{L_p}\right) \cosh\left(\frac{W_n - x}{L_p}\right). \end{aligned} \tag{3.10}$$

Evaluating at the depletion edge $x = 0$ yields the theoretical value of the diffusion current:

$$J_d(0) = -\frac{qp_{n0}L_p}{\tau_p} \operatorname{csch}\left(\frac{W_n}{L_p}\right). \tag{3.11}$$

Equation (3.11) is the well-known saturation current used in the ideal diode equation, capturing both the long- and short-base solutions. Figure 3·2 shows the minority carrier profiles and current densities as a function of the ratio of the quasi-neutral region width and minority carrier diffusion length. For small quasi-neutral regions and long diffusion lengths, the carrier profile is linear and the current density is constant with position and is solely dependent on the width of the n region, with shorter regions leading to higher dark current. For short diffusion lengths and long quasi-neutral regions the minority carrier concentration is able to return to its equilibrium value, and the dark current is independent of the n region length, and is dependent on the diffusion length.

3.1.2 Photocurrent density and quantum efficiency

Consider the case of uniformly illuminating the n side of the p-n junction, neglecting any surface reflection. Provided that the incident photon energy is greater than or equal to the bandgap of the semiconductor, the radiation is absorbed and generates electron-hole pairs. This can be included in (3.1) as a generation term modeled by the Lambert-Beer law where the intensity of the radiation falls exponentially with absorption depth for a single-pass of the radiation, or in terms of optical generation,

$$G_o(x) = \alpha\phi \mathrm{e}^{-\alpha(W_n - x)}, \tag{3.12}$$

where α and ϕ are the absorption coefficient and incident photon flux respectively. Under steady-state (3.1) becomes

$$D_p \frac{\partial^2 \Delta p}{\partial x^2} - \frac{\Delta p(x)}{\tau_p} = -\alpha\phi e^{-\alpha(W_n - x)}. \tag{3.13}$$

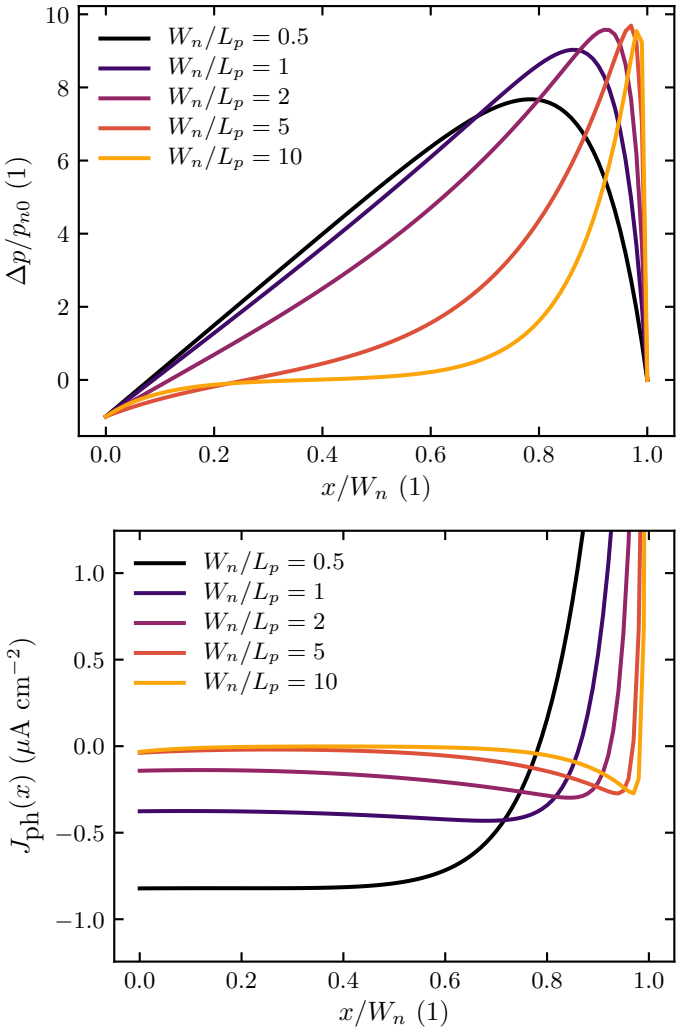

Figure 3·3: (Top) Excess minority carrier concentration and (bottom) current density under illumination for varied ratios of diffusion length and quasi-neutral region width. $\phi = 5 \cdot 10^{13}$ ph cm^{-2} s^{-2}, $\lambda = 1.5$ µm, and $\alpha = 8165.5$ cm^{-1}.

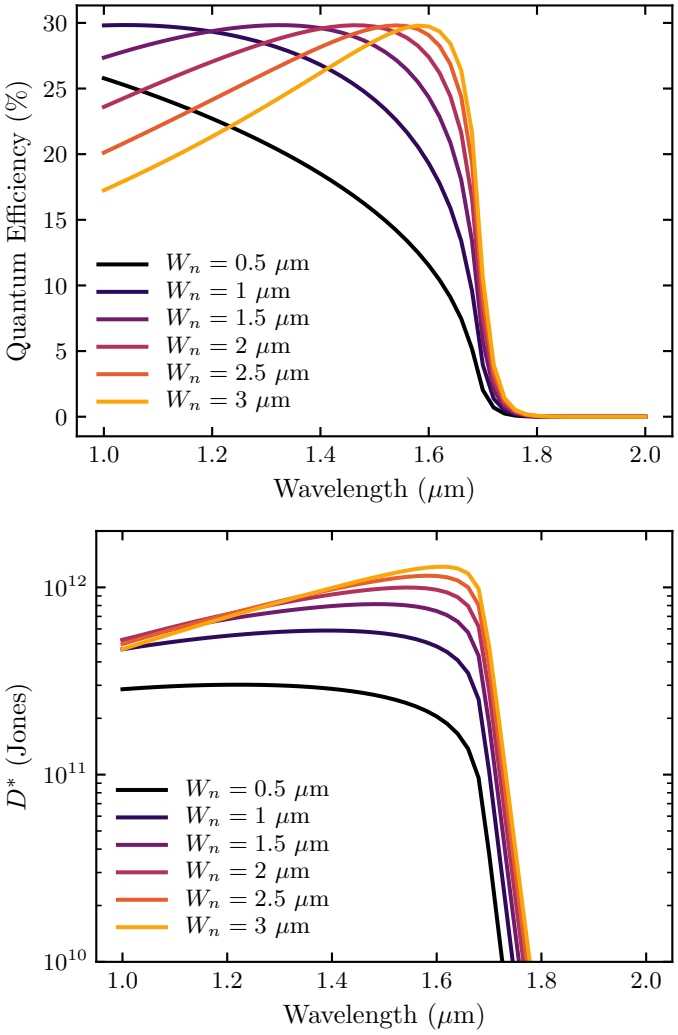

Figure 3·4: (Top) Diffusion-limited quantum efficiency and (bottom) specific detectivity for $In_{0.53}Ga_{0.47}As$ p-n junction at $T = 300K$ as a function of absorber thickness and wavelength.

The solution to this second-order inhomogenous differential equation is

$$\Delta p(x) = \gamma e^{-\alpha(W_n - x)} - \gamma e^{-x/L_p - W_n\alpha} - p_{n0} e^{-x/L_p} + 2\xi \sinh\left(\frac{x}{L_p}\right) \tag{3.14}$$

where

$$\gamma = \frac{-\alpha\phi\tau_p}{L_p^2\alpha^2 - 1},$$
$$\xi = \frac{1}{2}\left(\gamma e^{-\alpha W_n} - \gamma e^{W_n/L_p} + p_{n0}\right)\left[\coth\left(\frac{W_n}{L_p}\right) - 1\right]$$

The photocurrent density is then

$$J_{\mathrm{ph}}(x) = \\ -\frac{qL_p}{\tau_p}\left[\gamma\alpha L_p e^{-\alpha(W_n - x)} + \gamma e^{-x/L_p - W_n\alpha} + p_{n0} e^{-x/L_p} + 2\xi\cosh\left(\frac{x}{L_p}\right)\right], \tag{3.15}$$

and at the depletion edge

$$J_{\mathrm{ph}}(0) = \frac{-qL_p}{\tau_p}\left(\gamma\alpha L_p e^{-\alpha W_n} + \gamma e^{-\alpha W_n} + p_{n0} + \xi\right) \tag{3.16}$$

Figure 3·3 shows the resulting carrier profiles and current densities for $\lambda = 1.5$ µm, $\phi = 5{\cdot}10^{13}$ ph cm^{-2} s^{-1}, and $\alpha = 8165.5$ cm^{-1} for various ratios of diffusion length and quasi-neutral width. Due to the nearby contact maintaining the carrier concentration at its equilibrium value, the profile has a peak at a short depth into the device. The direct implication is shown in the current density where it has both positive and negative values, indicating a flux of optically generated carriers diffusing toward the cathode where they are collected and not sensed as additional photocurrent. For short devices or long diffusion lengths some carriers are able to be collected as photocurrent; for long devices the concentration returns to its equilibrium value and the only current is the original dark current.

We can quantify the lost photocurrent by looking at the quantum efficiency given

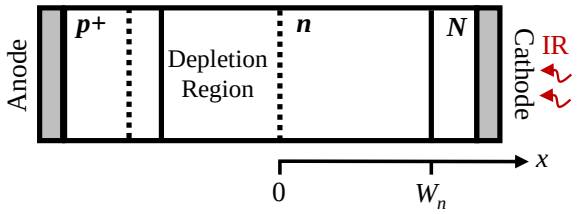

Figure 3·5: One-dimensional representation of a p-n-N junction.

by

$$\eta = \frac{J_{\text{ph}}(0) - J_d(0)}{q\phi}. \tag{3.17}$$

Generally, $In_{0.53}Ga_{0.47}As$, layers are thin, and the previous discussion would indicate that for the case of a p-n junction based device a thinner n region offers higher photocurrent. Figure 3·4 shows the expected QE and D^* from Eq. (1.6) for a p-n junction with $In_{0.53}Ga_{0.47}As$ materials parameters over wavelength. Exceptionally thin layers will have a lower QE for longer wavelengths as not enough of the radiation is absorbed before being transmitted. Large layers will have the QE decrease for shorter wavelengths as carriers are lost to the neighboring contact for shallow absorption depths. D^* includes the contribution from the dark current. Thinner layers will have a higher dark current and lower QE, and hence, D^* will be lower than the other cases. Considering all cases, the QE is lower than desired, with a peak value of about 30%.

3.2 Modified boundary conditions in planar focal plane arrays

The discussion of minority carrier current densities in the previous section is useful as an introduction to the continuity equations and solutions, and as an outline to estimating QE and D^*. However, in many FPAs the pixel architecture is not comprised solely of a p-n junction. Most infrared sensors contain several layers that are grown on a compatible substrate. The diode is subsequently formed in the subsequent processes.

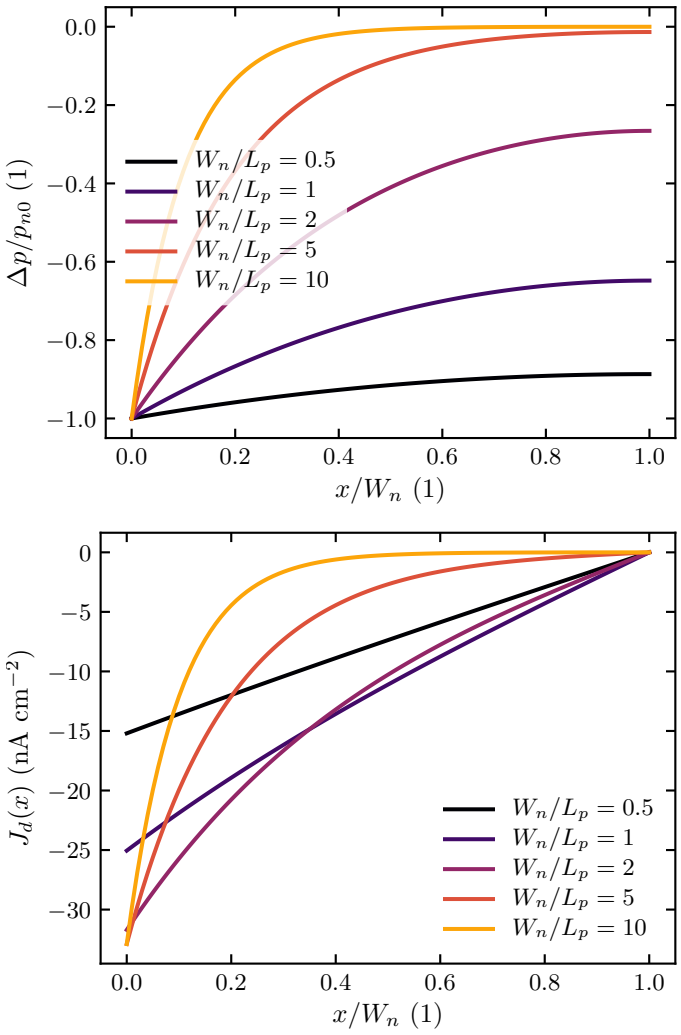

Figure 3·6: (Top) Normalized excess carrier density and (bottom) minority carrier current density versus normalized position for varied ratios of quasi-neutral width to diffusion length for a *p*-*n*-*N* junction with zero carrier flux boundary condition at $x = W_n$.

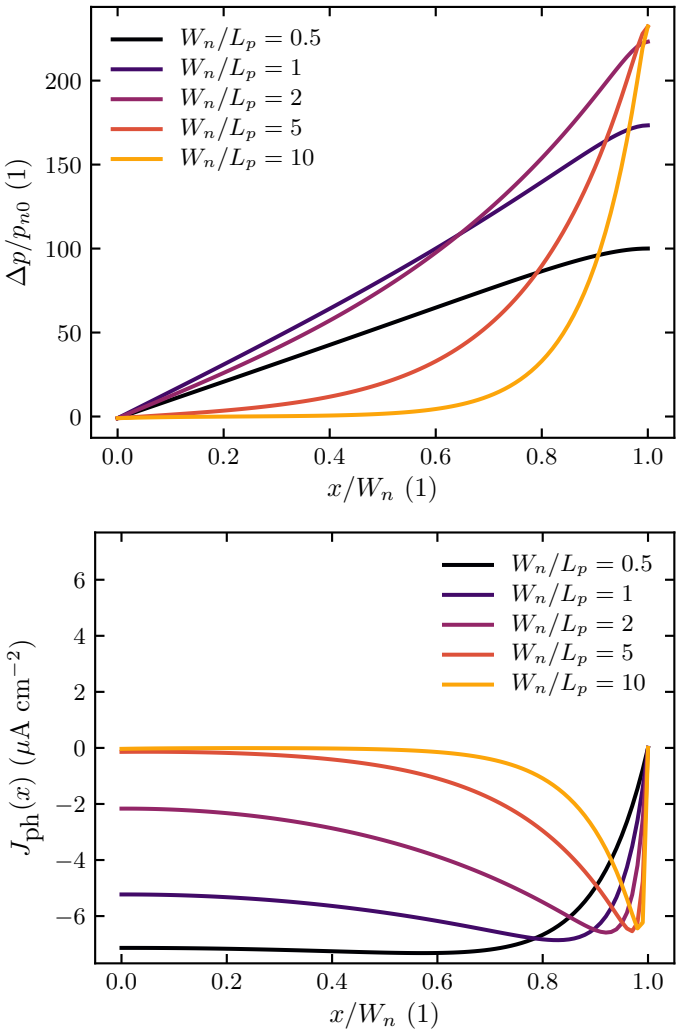

Figure 3·7: (Top) Normalized excess carrier density and (bottom) minority carrier density under illumination versus normalized position for varied ratios of quasi-neutral width and diffusion length for a p-n-N junction with zero carrier flux boundary condition at $x = W_n$. $\phi = 5 \cdot 10^{13}$ ph cm^{-2} s^{-2}, $\lambda = 1.5$ µm, and $\alpha = 8165.5$ cm^{-1}.

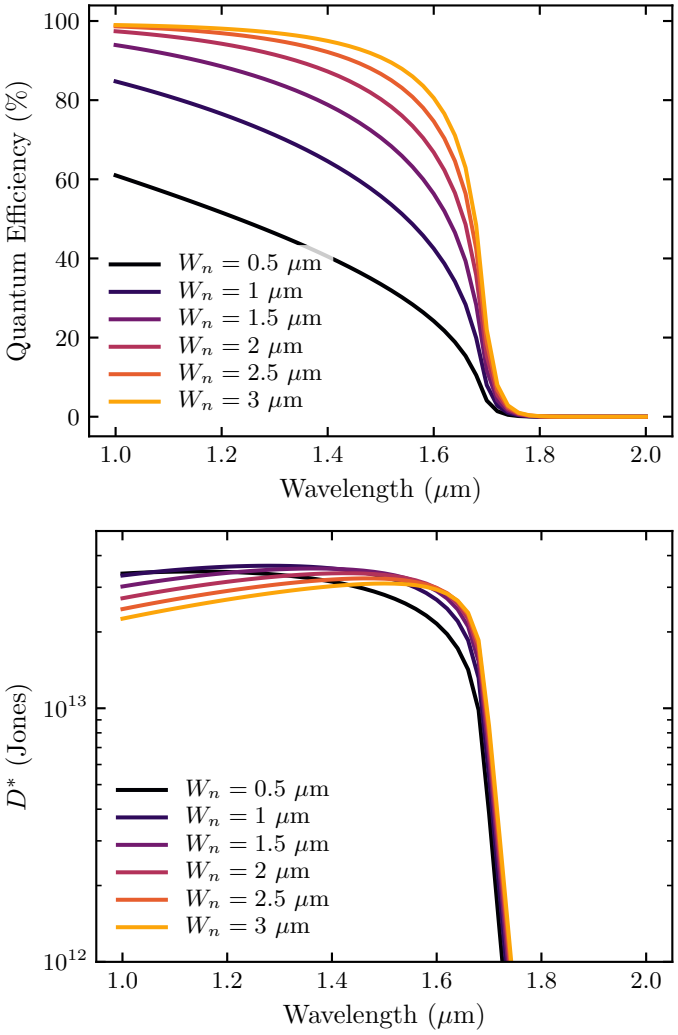

Figure 3·8: (Top) Diffusion limited quantum efficiency and (bottom) specific detectivity for an $In_{0.53}Ga_{0.47}As$ *p-n-N* based photodiode verus wavelength for varied absorber thickness.

Depending on the band alignment of the materials, this can change the boundary conditions used when considering the continuity equations. For $In_{0.53}Ga_{0.47}As$ DLPH photodiodes, the $In_{0.53}Ga_{0.47}As$ layers are grown on an InP substrate. Considering the case of an $N+$ substrate with an n-type absorber, there is a large valence band barrier to holes. This introduces a zero flux Neumann boundary condition. Vertically, we can study how this affects the minority carrier transport by applying the previous analysis to p-n-N junctions.

3.2.1 *p-n-N* junctions

The boundary conditions for the reference coordinate system in Fig. 3·5 are

$$\Delta p(0) = -p_{n0}, \tag{3.18}$$

$$\frac{\partial \Delta p(W_n)}{\partial x} = 0. \tag{3.19}$$

The solutions for $\Delta p(x)$, $J_d(x)$, and $J_d(0)$ are

$$\Delta p(x) = -p_{n0} \cosh\left(\frac{W_n - x}{L_p}\right) \operatorname{sech}\left(\frac{W_n}{L_p}\right), \tag{3.20}$$

$$J_d(x) = -\frac{qL_p p_{n0}}{\tau_p} \operatorname{sech}\left(\frac{W_n}{L_p}\right) \sinh\left(\frac{W_n - x}{L_p}\right), \tag{3.21}$$

$$J_d(0) = -\frac{qL_p p_{n0}}{\tau_p} \tanh\left(\frac{W_n}{L_p}\right). \tag{3.22}$$

The resulting excess carrier profile and current densities under dark conditions are shown in Fig. 3·6. For large quasi-neutral regions or short diffusion lengths the minority carrier profile is able to return to its equilibrium values. For short regions or long diffusion lengths this is no longer the case; the minority carrier profile is suppressed below its equilibrium value through the entire region. This manifests as a reduction in the minority carrier current density. In contrast to the original p-n junction where a long diffusion length with a short quasi-neutral region resulted in a

larger current density due to the steeper minority carrier slope as the contact fixes the minority carrier density at its thermal equilibrium value, the *p*-*n*-*N* junction results in a much lower current density.

Again assuming a Beer's law absorption, and assuming the N region has a bandgap large enough to not absorb any of the incident radiation, the minority carrier concentration and current densities under illumination have the same forms as (3.14) and (3.15) with

$$\xi = -\frac{\gamma e^{-\alpha W_n} + \gamma\alpha L_p e^{W_n/L_p} + p_{n0}}{e^{2W_n/L_p} + 1}. \tag{3.23}$$

Figure 3·7 presents the modified excess carrier profiles and current densities under illumination. Now, instead of having a peak excess carrier density away from the quasi-neutral region edge, the peak is at exactly $x = W_n$. Consequently the current density is always negative with position, indicating carriers diffusing towards the depletion edge; no carriers are lost by diffusing away from the depleted edge, like what was observed in the *p*-*n* case. However, if the quasi-neutral region is too long, or diffusion length too short, the carriers will recombine before reaching the depletion edge and will not be sensed as additional current, like the previous *p*-*n* case. These behaviors are also reflected in Fig. 3·8 where the QE is nearly perfect for thicker quasi-neutral regions and D^* is improved by an order of magnitude, due to the higher QE and lower dark current.

3.2.2 *pnp*-junctions

In a planar array of photodiodes the neighboring junctions may interact with each other and change the minority carrier profiles laterally through the device. A useful way to study this is to consider the case of a *p*-*n*-*p* junction shown in Fig. 3·9. In this case, the boundary conditions are

$$\Delta p(0) = -p_{n0}, \tag{3.24}$$

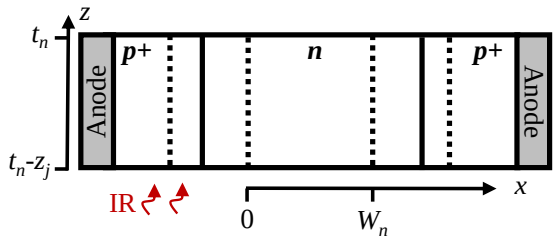

Figure 3·9: One-dimensional representation of two neighboring photodiodes comprising a *p*-*n*-*p* junction. In this case, the optical generation is constant with respect to x and depends on z and the height of the *p*-*n*-*p* junction with respect to the illuminated surface—in this example the base of the junction is illuminated.

$$\Delta p(W_n) = -p_{n0}, \tag{3.25}$$

assuming both diodes are held had the same sufficiently large reverse bias. With this set of symmetric boundary conditions, the solutions to the continuity equation under steady-state are

$$\Delta p(x) = -p_{n0}\,\mathrm{sech}\left(\frac{W_n}{2L_p}\right)\cosh\left(\frac{W_n - 2x}{2L_p}\right), \tag{3.26}$$

$$J_d(x) = -\frac{qp_{n0}L_p}{\tau_p}\,\mathrm{sech}\left(\frac{W_n}{2L_p}\right)\sinh\left(\frac{W_n - 2x}{2L_p}\right), \tag{3.27}$$

$$J_d(0) = -\frac{qp_{n0}L_p}{\tau_p}\tanh\left(\frac{W_n}{2L_p}\right). \tag{3.28}$$

Notice that the above solutions are identical to the previous solutions to the *p*-*n*-*N* junction if the right-hand-side boundary was located at $x = W_n/2$. The implication is that when considering two neighboring junctions there is a symmetry plane of zero flux half way between the two junctions where the minority carrier profile reaches a maximum value, as shown in Fig. 3·11. The dark current density has both positive and negative values, indicating carriers diffusing from the center towards each junction. For large distances between neighboring junctions, or equivalently short diffusion lengths, the junctions behave as normal long-base diodes where the dark current

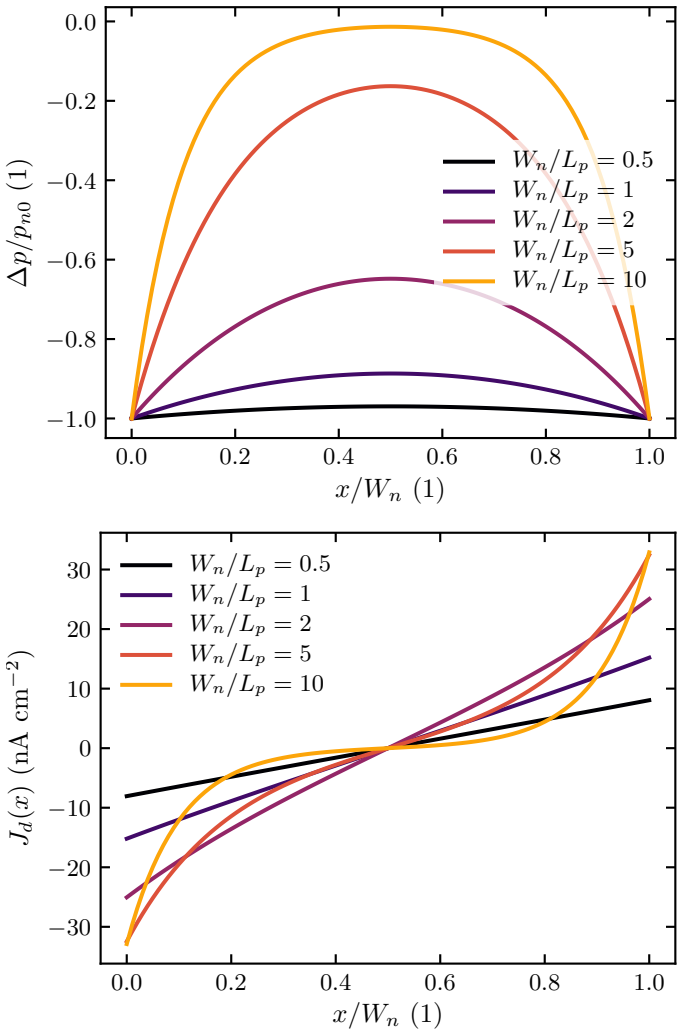

Figure 3·10: (Top) Normalized excess carrier density and (bottom) minority carrier current density versus normalized position for varied ratios of quasi-neutral width to diffusion length for a *p*-*n*-*p* junction with zero carrier flux boundary condition at $x = W_n$.

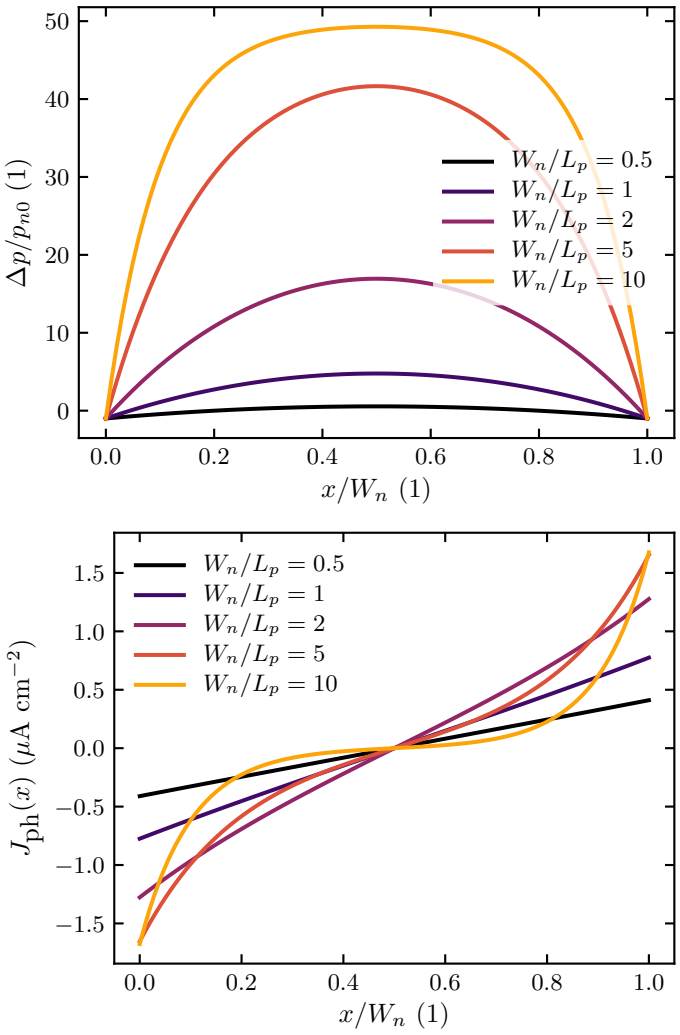

Figure 3·11: (Top) Normalized excess carrier density and (bottom) minority carrier density under illumination versus normalized position for varied ratios of quasi-neutral width and diffusion length for a *p*-*n*-*p* junction. $\phi = 5 \cdot 10^{11}$ ph cm^{-2} s^{-2}, $\lambda = 1.5$ µm, and $\alpha = 8165.5$ cm^{-1}.

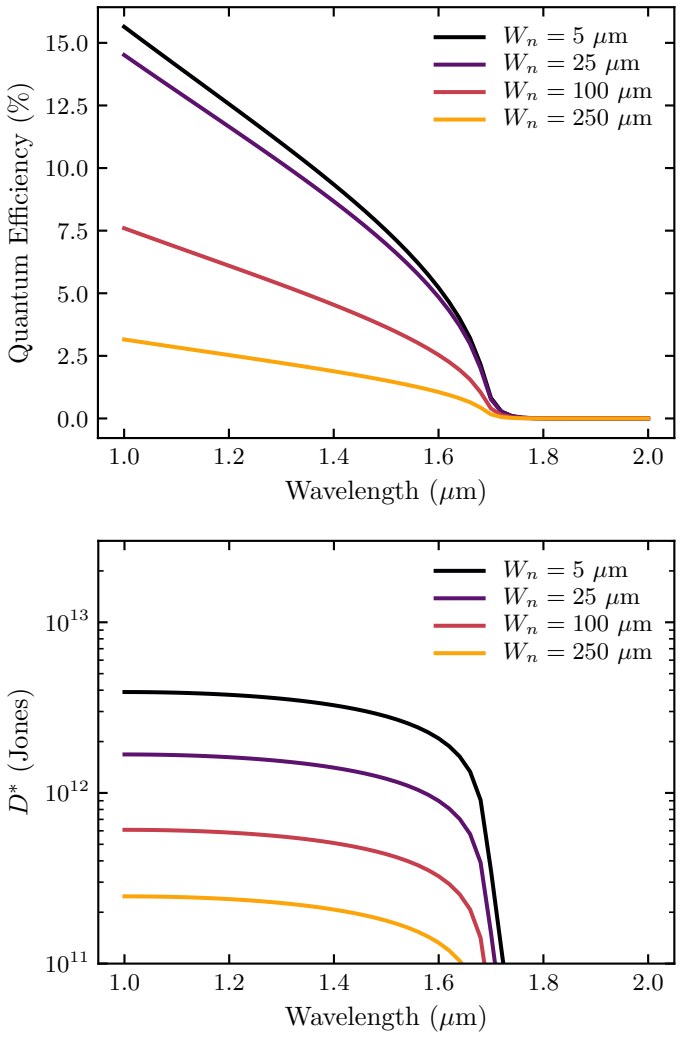

Figure 3·12: (Top) Diffusion limited quantum efficiency and (bottom) specific detectivity for an $In_{0.53}Ga_{0.47}As$ *p-n-p* based photodiode verus wavelength for varied distance between *p-n* junctions. The thickness along z is assumed to be 200 nm.

depends on the minority carrier diffusion length. For long diffusion lengths, or short distances between junctions, the minority carrier is substantially suppressed; the dark current falls steadily with shrinking distance.

The direction of illumination is orthogonal to x in this discussion, that is, the optical generation rate is determined by the depth of the p-n-p junction and is dependent on z and constant with x. It can be shown that

$$\Delta p(x) = G(z)\tau_p - (p_{n0} + G(z)\tau_p)\,\mathrm{sech}\left(\frac{W_n}{2L_p}\right)\cosh\left(\frac{W_n - 2x}{2L_p}\right), \tag{3.29}$$

where

$$G(z) = \alpha\phi e^{-\alpha z}. \tag{3.30}$$

The photocurrent density is then given by

$$J_p(x) = -\frac{qL_p}{\tau_p}\left[(p_{n0} + G(z)\tau_p)\,\mathrm{sech}\left(\frac{W_n}{2L_p}\right)\sinh\left(\frac{W_n - 2x}{2L_p}\right)\right], \tag{3.31}$$

$$J_p(0) = -\frac{qL_p}{\tau_p}\left[(p_{n0} + G(z)\tau_p)\tanh\left(\frac{W_n}{2L_p}\right)\right]. \tag{3.32}$$

The minority carrier profile and current density under illumination has the same forms as under dark conditions, as shown in Fig. 3·11. With enough distance between the junctions, the excess minority carrier concentration will reach a constant value determined by the product of the optical generation rate and minority carrier lifetime.

In the previous discussions of p-n and p-n-N junctions the illumination direction was the same as the current density. Hence, the areas for the current density and photon flux in (3.17) would cancel without needing to convert the quantities to carriers per second and incoming photons per second. However, in this case the photon flux is incident in the xy-plane while the current density along x is into the yz-plane. Therefore to accurately calculate the QE it is necessary to compute the total current by integrating $J_d(0)$ and $J_{\mathrm{ph}}(0)$ along z from $z = t_n - z_j$ to $z = t_n$ where t_n is the

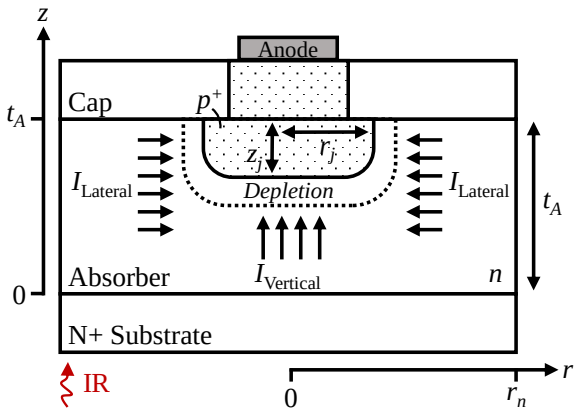

Figure 3·13: Schematic of a DLPH photodiode highlighting the sources of diffusion current.

thickness along z of the absorbing region, and z_j is the junction depth measured from the top of the layer, and multiply ϕ by W_n. The result is shown in Fig. 3·12. The low QE is due to the assumed thickness of the *p-n-p* junction of only 200 nm, and that in a true representation of the array, the QE would be two times greater; the exemplar *p-n-p* junction would be mirrored on the left-hand-side leading to twice the photocurrent. We would use the same value of W_n to normalize ϕ since this would be the effective optical input area of the pixel of interest. Finally, it is observed that the QE approaches its peak values for each wavelength as long as the junctions are within a diffusion length of each other, and does not degrade as the inter-pixel spacing is reduced.

3.2.3 Lateral versus vertical transport

Using the insight from the previous one-dimensional analysis, we can develop an approximation for the dark current and photocurrent for a three-dimensional pixel in a planar photodiode array. One approach is to solve the continuity equations in either spherical or cylindrical coordinates, with appropriate boundary conditions (Grimbergen, 1976, Trezza et al., 2011, Wichman et al., 2014). Assuming cylindrical coor-

dinates (2.32) in the quasi-neutral regions of the device the continuity equation for holes becomes

$$\begin{aligned}\frac{\partial \Delta p}{\partial t} &= D_p \nabla^2 \Delta p - \frac{\Delta p}{\tau_p} \\ &= D_p \left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \Delta p}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \Delta p}{\partial \theta^2} + \frac{\partial^2 \Delta p}{\partial z^2} \right] - \frac{\Delta p}{\tau_p}. \end{aligned} \tag{3.33}$$

As shown in Fig. 3·13, we assume that there are two components that contribute to the current. The first is a lateral current that depends on the pixel dimensions, and the second is a vertical component that is determined by the thickness of the material and junction radius and depth. Assuming no azimuthal or axial dependence, the lateral component only depends on the radial distance, r. Laterally, (3.33) becomes under steady-state and dark conditions

$$\frac{\partial^2 \Delta p}{\partial r^2} + \frac{1}{r} \frac{\partial \Delta p}{\partial r} - \frac{\Delta p}{L_p^2} = 0, \tag{3.34}$$

whose solutions are modified Bessel functions of order zero. From the previous discussions of *p-n-p* junctions we can reasonably assume a zero flux boundary condition on the pixel edges[1]. The appropriate boundary conditions are

$$\Delta p(r_0) = -p_{n0}, \tag{3.35}$$

$$\frac{\partial \Delta p(r_n)}{\partial r} = 0, \tag{3.36}$$

where $r_0 = r_j + W_d$ and the quasi-neutral width r_n depends on θ as $r_n(\theta) = W \sec(\theta)/2$ for $\theta = 0$ to $\pi/4$ for the case of a square pixel of pitch W. W_d is the depletion width[2].

[1]Wichman et al. also showed with three-dimensional numerical simulation results that there are symmetry lines along pixel boundaries in rectangular arrays where the minority carrier profile reaches a maximum, and noted that this can be interpreted as a zero flux condition along the pixel edges (Wichman et al., 2014).

[2]For this analysis we assume W_d is the one-dimensional width. Recall that for an asymmetric junction with $N_a \gg N_d$, W_d can be approximated by $\sqrt{2\varepsilon(V_{bi} - V_a)/qN_d}$ where V_{bi} is the built-in potential given by $kT \ln(N_d N_a / n_i^2)/q$ (Muller and Kamins, 2003).

The solutions for Δp and J_d from (3.40) with these boundary conditions are

$$\Delta p(r,\theta) = -p_{n0}\left(\frac{I_1(r_n(\theta)/L_p)K_0(r/L_p) + I_0(r/L_p)K_1(r_n(\theta)/L_p)}{I_1(r_n(\theta)/L_p)K_0(r_0/L_p) + I_0(r_0/L_p)K_1(r_n(\theta)/L_p)}\right), \tag{3.37}$$

$$J_l(r,\theta) = \frac{-qL_p p_{n0}}{\tau_p}\left(\frac{I_1(r_n(\theta)/L_p)K_1(r/L_p) - I_1(r/L_p)K_1(r_n(\theta)/L_p)}{I_1(r_n(\theta)/L_p)K_0(r_0/L_p) + I_0(r_0/L_p)K_1(r_n(\theta)/L_p)}\right), \tag{3.38}$$

where $I_\alpha(x)$ and $K_\alpha(x)$ are the modified Bessel functions of the first and second kind respectively. To calculate the total current due to the lateral component, we integrate $J_l(r_0,\theta)$ along a cylindrical surface of thickness, t_A:[3]

$$\begin{aligned} I_l &= -\frac{q8L_p p_{n0} r_0}{\tau_p}\int_0^{\pi/4}\int_0^{t_A} J_l(r_0,\theta)dzd\theta \\ &= -\frac{q8L_p p_{n0} r_0 t_A}{\tau_p}\int_0^{\pi/4} J_l(r_0,\theta)d\theta. \end{aligned} \tag{3.39}$$

The integral in (3.39) does not have a closed-form expression for the case of a square pixel, and must be solved numerically[4].

While illuminated the differential equation becomes

$$\frac{\partial^2\Delta p}{\partial r^2} + \frac{1}{r}\frac{\partial\Delta p}{\partial r} - \frac{\Delta p}{L_p^2} = -\frac{G(z)\tau_p}{L_p^2}, \tag{3.40}$$

where

$$G(z) = \alpha\phi e^{-\alpha z}.$$

[3]A more physically appropriate integration surface would be along the depleted cylindrical surface. This, however, neglects an important contribution to the current from the bottom corners of the pixel. To approximate these corner effects we extend the integration surface through the whole device.

[4]If we neglect the azimuthal dependence of the quasi-neutral width, the surface integrals become trivial. However, this causes the analytical model to systematically underestimate the total current due to the currents from the corners since the model does not capture the higher current densities from the larger distance from the junction to the pixel corners.

The resulting photocurrent density at $r = r_0$ is

$$J_{l,ph} = J_l(r_0, \theta) \left(1 + \frac{G(z)\tau_p)}{p_{n0}} \right). \tag{3.41}$$

The total current is given by

$$I_{l,ph} = I_l \left(1 + \frac{\tau_p \Delta G}{p_{n0} \alpha t_A} \right), \tag{3.42}$$

where $\Delta G = G(0) - G(t_A)$.

Without radial and azimuthal dependence the vertical current component reduces to the same one-dimensional differential equations used in the previous discussion of *p-n-N* junctions. The total currents are obtained by simply integrating along the circular surface at the bottom of the cylindrical depleted region. Hence, the total vertical current is just the current densities multiplied by $\pi(r_j + W_d)^2$.

To validate the model we compare the expected dark current and quantum efficiency with a 3D drift-diffusion simulation of a similar device. More details about the numerical model are discussed in Chapter 4. Shown in Fig. 3·14, the analytical model shows excellent agreement with representative drift-diffusion simulations.

Figure 3·15 shows the lateral and vertical dark current components as a function of the ratio of pixel pitch to diffusion length. Once the pixel pitch approaches the minority carrier diffusion length, the total dark current falls with reducing pitch. The primary source of dark current, assuming r_j is small compared to the pitch, is lateral diffusion from the pixel boundaries. Once the pitch is reduced well below the diffusion length, the vertical component becomes important. Of course, if r_j is large, and if the material is thicker like what is commonly used for MW/LW detectors, the vertical component may be dominant.

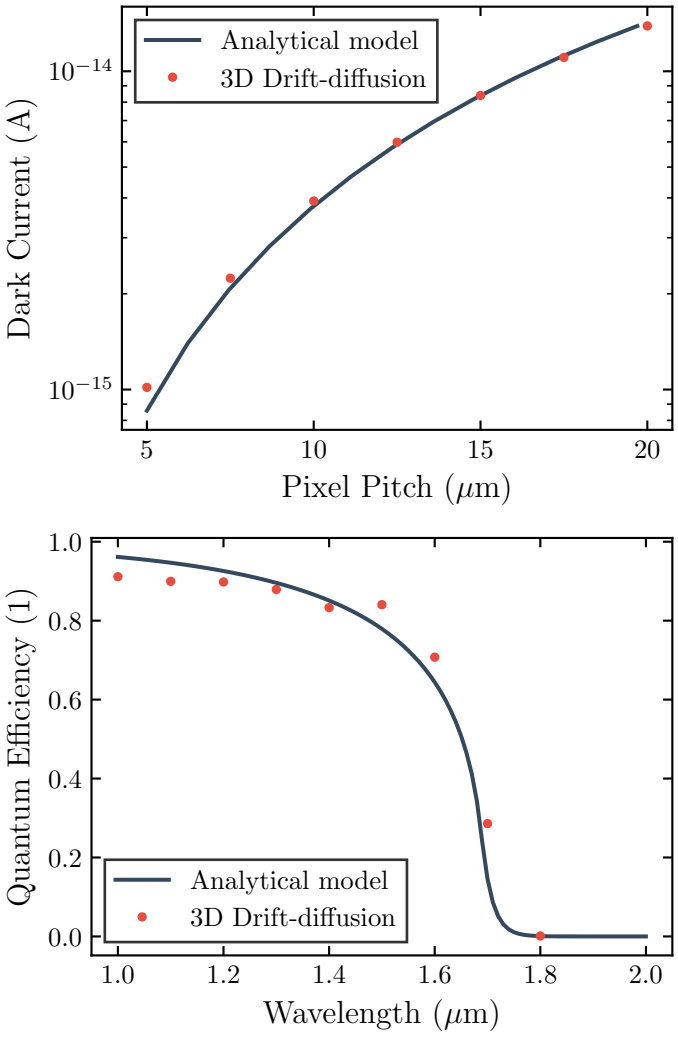

Figure 3·14: (Top) A comparison of the predicted dark current by the pseudo-3D analytical model with drift-diffusion simulation as a function of pixel pitch. (Bottom) A comparison of the analytical quantum efficiency with a numerical simulation versus wavelength. The simulated QE has been corrected to neglect surface reflection from the InP substrate to align with the assumptions in the analytical modeling.

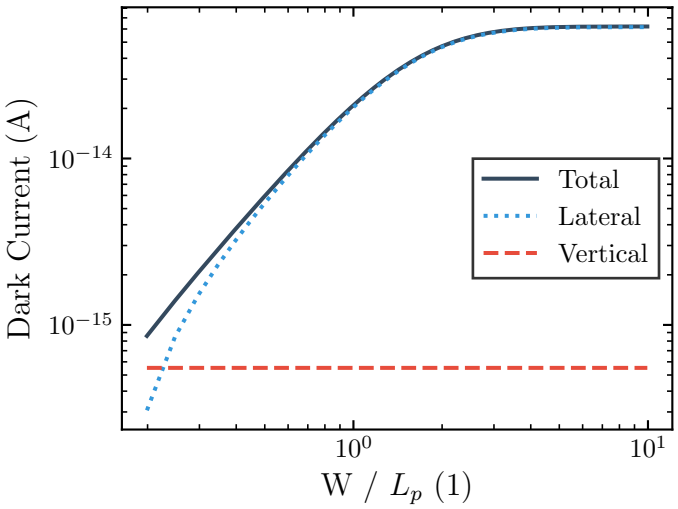

Figure 3·15: (Right) Dark current versus ratio of pixel pitch and diffusion length for the vertical and lateral current components. $r_j = 2$ µm, $t_A = 3$ µm, $V_a = -200$mV, $T = 300$K.

3.3 Conclusion

In this chapter we explored the solutions to the minority carrier continuity equations to derive equations for the diffusion-limited dark current, QE, and D^* in *p-n*, *p-n-N*, and *p-n-p* junctions. It was shown that due to the zero flux boundary condition for a *p-n-N* heterojunction reflecting minority carriers away from the cathode, the QE is improved from 30% to almost perfect when compared to a *p-n* homojunction. It was also noted that once the distance between two neighboring diodes is reduced to within a diffusion length, the minority carrier profile is suppressed, resulting in a lower current density. Finally, by extending the analysis to cylindrical coordinates a model for the dark current and QE in a three-dimensional photodiode was developed and validated by a three-dimensional drift-diffusion simulation. From this, it was shown that a substantial contribution to the dark current is from the lateral diffusion of minority carriers from the pixel boundaries.

Chapter 4

Dense array effects in planar arrays

This chapter reviews the consequences of moving towards smaller pixel pitches in planar infrared focal plane arrays. Though the chapter focuses on $In_{0.53}Ga_{0.47}As$ FPAs for SWIR sensing, much of the discussion is applicable to other detectors and spectral bands. First, we introduce the numerical model and compare the predicted dark current and quantum efficiency of a typical pixel to data from literature. Then, we explore how the dark current, QE, and MTF change as the pixel pitch is reduced.

4.1 $In_{0.53}Ga_{0.47}As$ Focal Plane Array Numerical Model

The methodology for simulating infrared focal plane arrays in this dissertation was adapted from previous works (Wichman et al., 2014, Pinkie et al., 2013, Bellotti and D'Orsogna, 2006). Figure 4·1 presents an example of a 2x2 DLPH pixel array. The epitaxial stack consists of an N^+ InP substrate, an n $In_{0.53}Ga_{0.47}As$ absorber, and an N InP capping layer. We assume that the p^+ diffusion forms an abrupt cylindrical junction that persists 200 nm into the absorber. Unless otherwise noted, the operating temperature was fixed at 300K and a applied bias of -200 mV was applied to each pixel anode. Based on other work, the SRH lifetime was set to 107 μs (DeWames et al., 2015) with a defect level at the intrinsic Fermi energy. The p^+ and n-type doping densities were 10^{18} cm^{-3} and 10^{16} cm^{-3} respectively.

The model was validated against data available in literature for similar FPA ar-

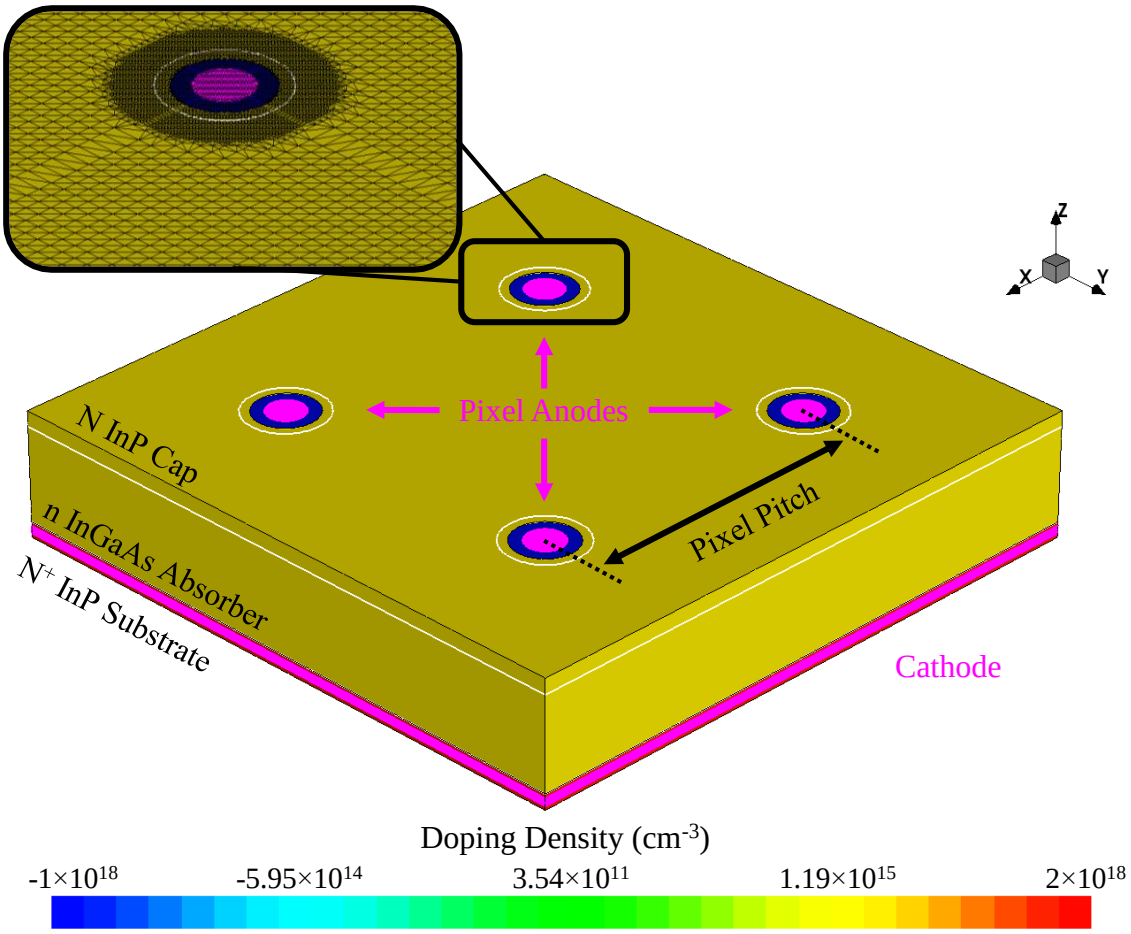

Figure 4·1: A finite-element mesh of a 2x2 $In_{0.53}Ga_{0.47}As$ focal plane array with an inset highlighting the mesh discretization strategy.

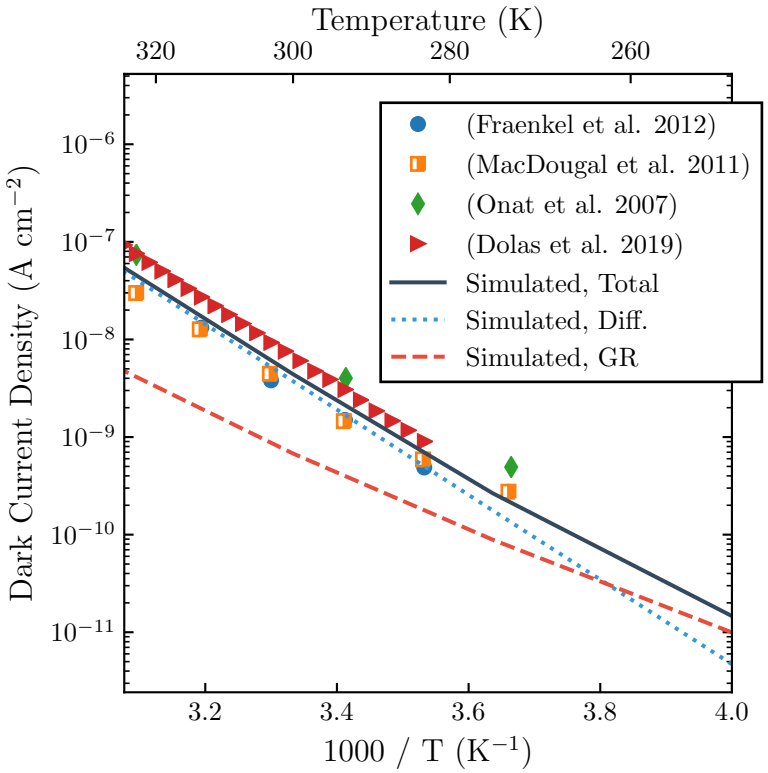

Figure 4·2: A comparison of the simulated dark current density of a SWIR $In_{0.53}Ga_{0.47}As$ pixel to various sources in literature (Onat et al., 2007, MacDougal et al., 2011, Fraenkel et al., 2012, Dolas et al., 2019). The simulation parameters include a p^+ doping density of 10^{18} cm^{-3}, n-type doping density of 10^{16} cm^{-3}, absorber thickness of 3 µm, $V_a = -200$ mV, $\tau_{\text{SRH}} = 107$ μs.

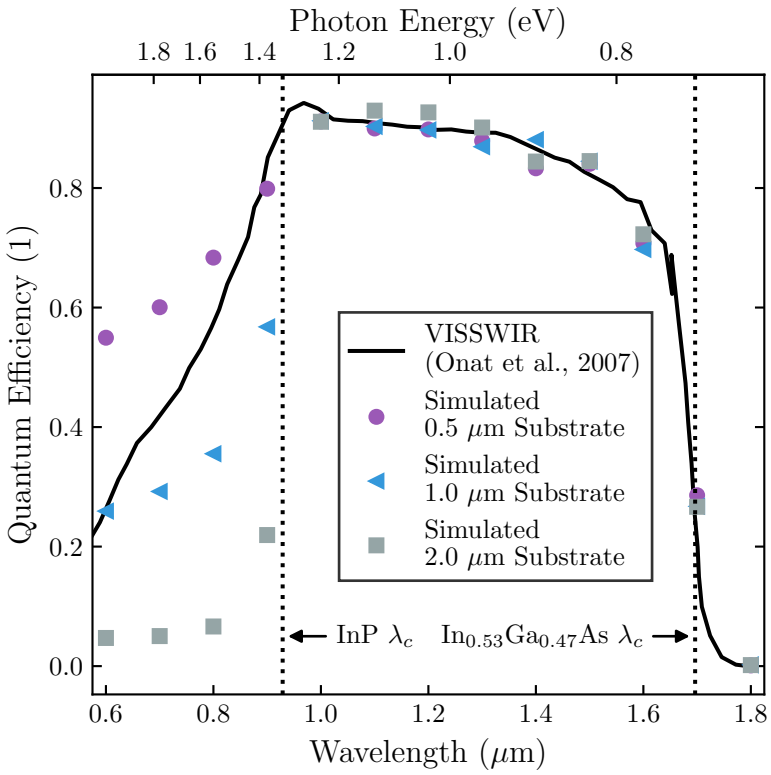

Figure 4·3: A comparison of the simulated QE with data from literature with a thinned InP substrate to extend photoresponse further into the visible range (Onat et al., 2007).

chitectures[1]. The dark current, shown in the Arrhenius plot in Fig. 4·2, matches well with several sources. In particular, the device is diffusion-limited to about 260K with a slope consistent with experimental data. The discrepancies in the y-intercept can be attributed to minor differences in doping density or absorber thickness which lead to a higher or lower saturation current.

We also compared the model's performance against experimental QE data, shown in Fig. 4·3. The InP substrates are commonly thinned, or removed entirely, on $In_{0.53}Ga_{0.47}As$ focal plane arrays to extend the detector's response deeper into the visible spectrum (Onat et al., 2007, Yuan et al., 2012). By thinning the substrate, less radiation is absorbed by the substrate for photon energies greater than the bandgap of InP. By reducing the substrate thickness from 2 µm to 0.5 µm the QE at $E_{ph} = 1.4$ eV rises from 20% to close to 80%.

4.2 Dark current

In the previous chapter it was noted that if two neighboring photodiodes are within two diffusion lengths of each other, the minority carrier profile laterally will be suppressed below its equilibrium value, resulting in a reduced lateral current. Figure 4·4 validates this effect, where the hole density is plotted laterally through the center of two photodiodes just below the depletion region edge. The diffusion length of $In_{0.53}Ga_{0.47}As$ for this device is around 25 µm, meaning the profile is suppressed for the largest pixel pitch simulated as well. The total dark current per pixel is also shown in Fig. 4·4. From 20 µm to 5 µm the total dark current drops by an order of magnitude.

By running simulations with and without SRH recombination we can isolate the diffusion and GR contributions to the dark current. Shown in Fig. 4·4 the reduction in

[1]The exact pixel architecture—layer thicknesses and doping densities—are often not reported in literature.

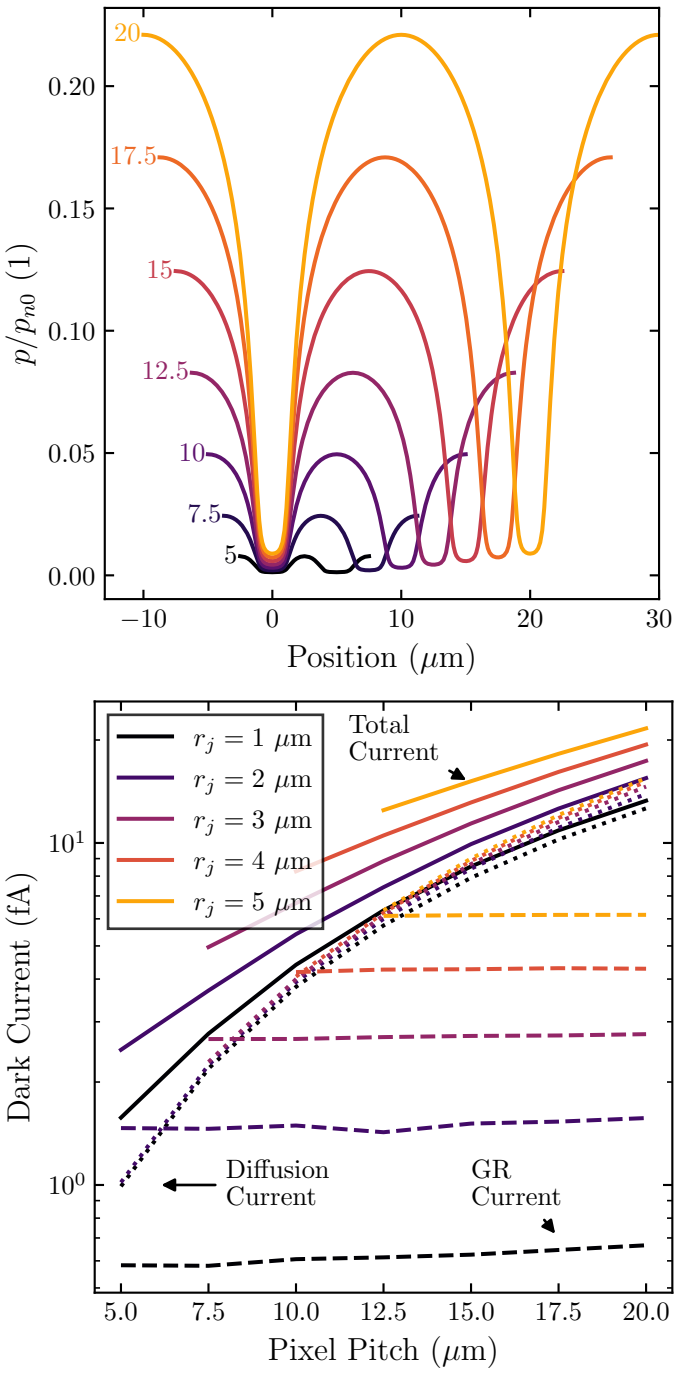

Figure 4·4: (Top) Normalized hole density laterally through the center of two adjacent three-dimensional pixels 100 nm below the depletion region edge for varied pixel pitch. (Bottom) Total dark current as a function of pixel pitch and junction radius. Dashed and dotted lines represent the GR and diffusion contributions to the current.

dark current is due to a decrease in the diffusion current, consistent with a reduction in lateral diffusion currents as the neighboring photodiodes interact to a higher degree. The GR current is pitch invariant;[2] generation-recombination current depends on the volume of the depleted region, which for constant r_j, t_A, and z_j is unchanged.[3]

4.3 Quantum efficiency and D^*

The quantum efficiency and specific detectivity of an $In_{0.53}Ga_{0.47}As$ pixel were simulated for varied pixel pitch as shown in Fig. 4·5. The QE rises by about 10% from 20 µm to 5 µm pitch. This is due to how the study was performed; each case had the same junction radius of 2 µm. The implication is that for larger pitches there is a larger quasi-neutral region that the carriers must diffuse through before reaching the junction. During the long diffusion path there is a higher likelihood of carriers recombining, leading to a lower number of carriers per second reaching the junction when compared to the total number of photon per second incident on the device.

The specific detectivity is relatively unchanged with pitch, decreasing by about a factor of two as the pitch is reduced. While the total dark current is lower for a smaller pitch, the dark current density is higher as shown in Fig. 4·6. As the pitch is reduced and the lateral contribution to the dark current is suppressed, the dark current is limited by GR from the depletion region and the vertical current originating from directly below the junction, which for constant junction radius is unchanged. As such, the dark current density rises with smaller pitches. The QE does not improve enough to mitigate the higher dark current density, so D^* goes down slightly.

Lastly, QE and D^* for decreasing absorber thickness are also shown in Fig. 4·5. For larger pixel pitches a thinner absorber leads to higher D^*; though reducing the absorber thickness from 3 µm to 1 µm results in 15% QE loss, the dark current density

[2]The slight variations in GR current is due to the mesh discretization.

[3]The GR current depends only on t_A when the depletion region extends through the whole layer.

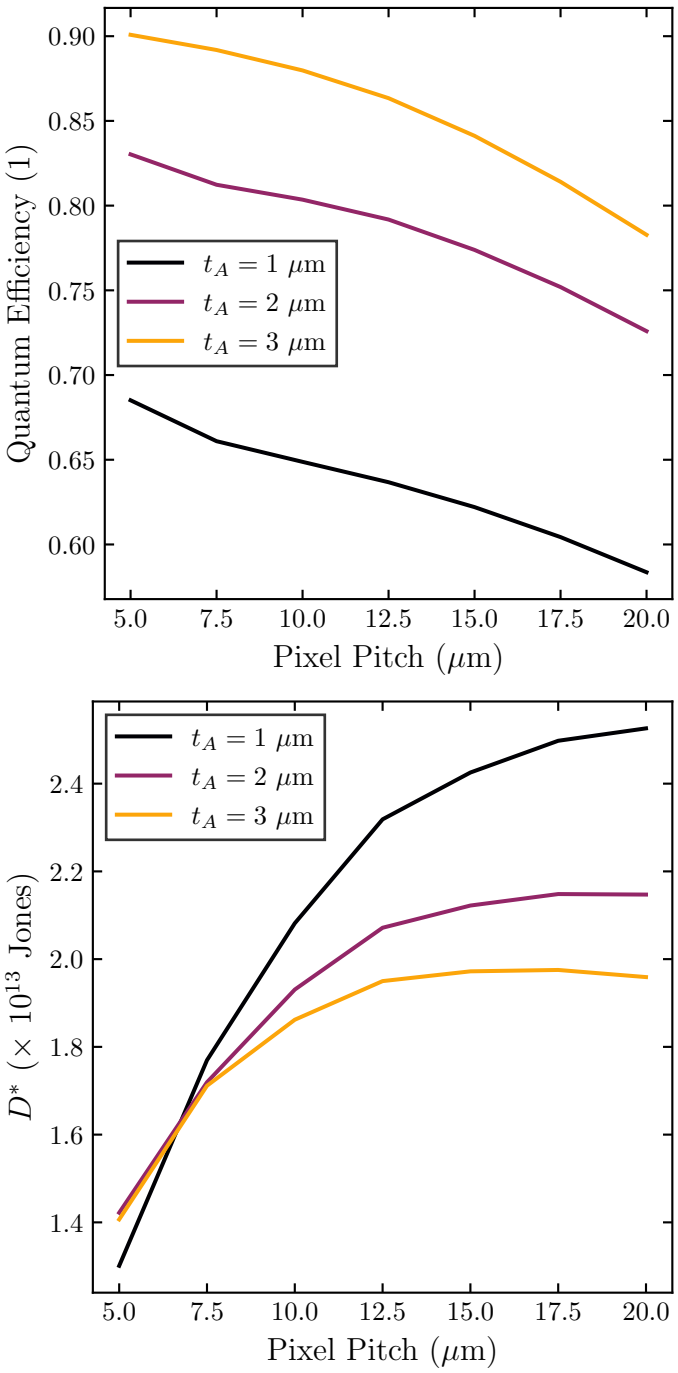

Figure 4·5: (Top) QE versus pitch and absorber thickness. The reported QE values have been corrected to account for reflections off of the InP substrate, similar to an ideal antireflection coating. (Bottom) Specific detectivity versus pitch and absorber thickness. $T = 300$K, $\lambda = 1.5$ µm, $V_a = -200$mV, $r_j = 2$ µm.

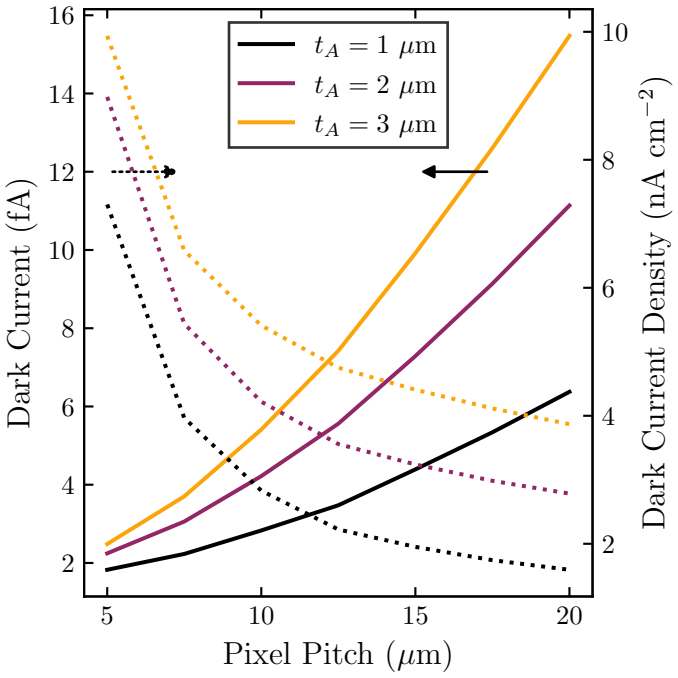

Figure 4·6: Dark current and dark current density versus pixel pitch. Solid lines represent total dark current on the left axis; dotted lines represent dark current density on the right axis.

is also reduced by more than a factor of two. As the pixel pitch is reduced this effect is reduced, since the dark current density rises as the pixel is dominated by the GR and vertical diffusion currents. Overall the relative change in D^* is marginal, lowering by less than a factor of two over the range of pixel sizes considered in this work.

4.4 Modulation transfer function

One advantage of moving towards a smaller pixel pitch is that the detector's modulation transfer function should improve. There are, however, potential challenges in realizing this improvement. For example, there is inherently more inter-pixel crosstalk as the pixels are closer together. Using the methodology outlined in Chapter 2, we simulate the MTF by taking the Fourier transform of a spot-scan profile of photocurrent versus Gaussian beam position, shown in Fig. 4·7. The simulation consisted of

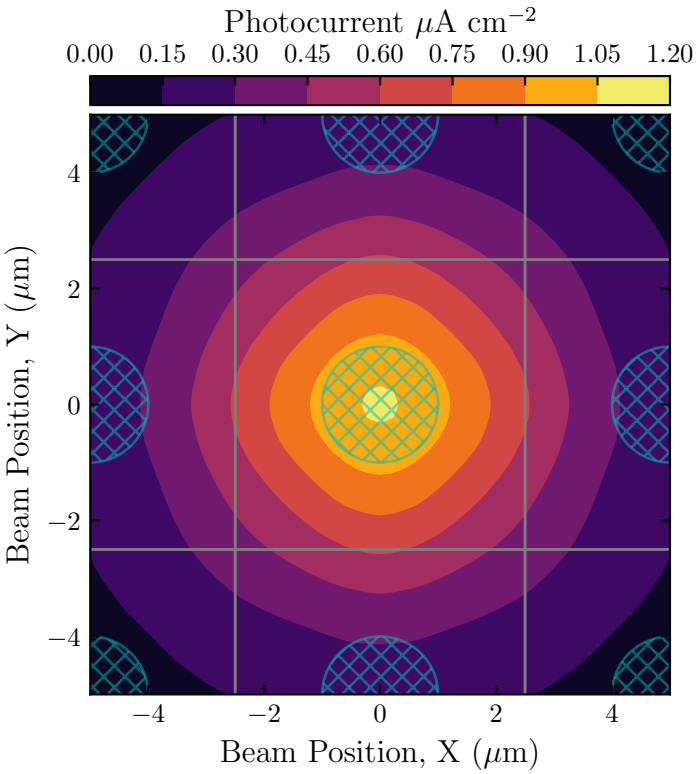

Figure 4·7: Two-dimensional spot-scan profile for a 5 µm pitch focal plane array. Hatched circles are used to represent the p^+ diffusion and the gray lines indicate the pixel borders.

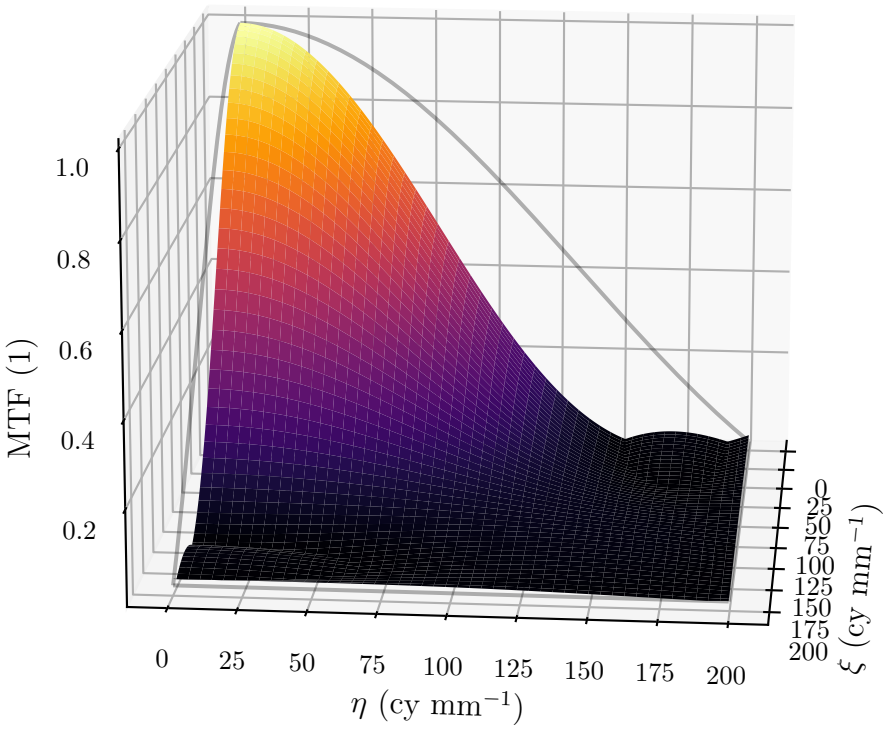

Figure 4·8: Full modulation transfer function in terms of both spatial frequency components. The dark gray border is the detector footprint considering a square pixel.

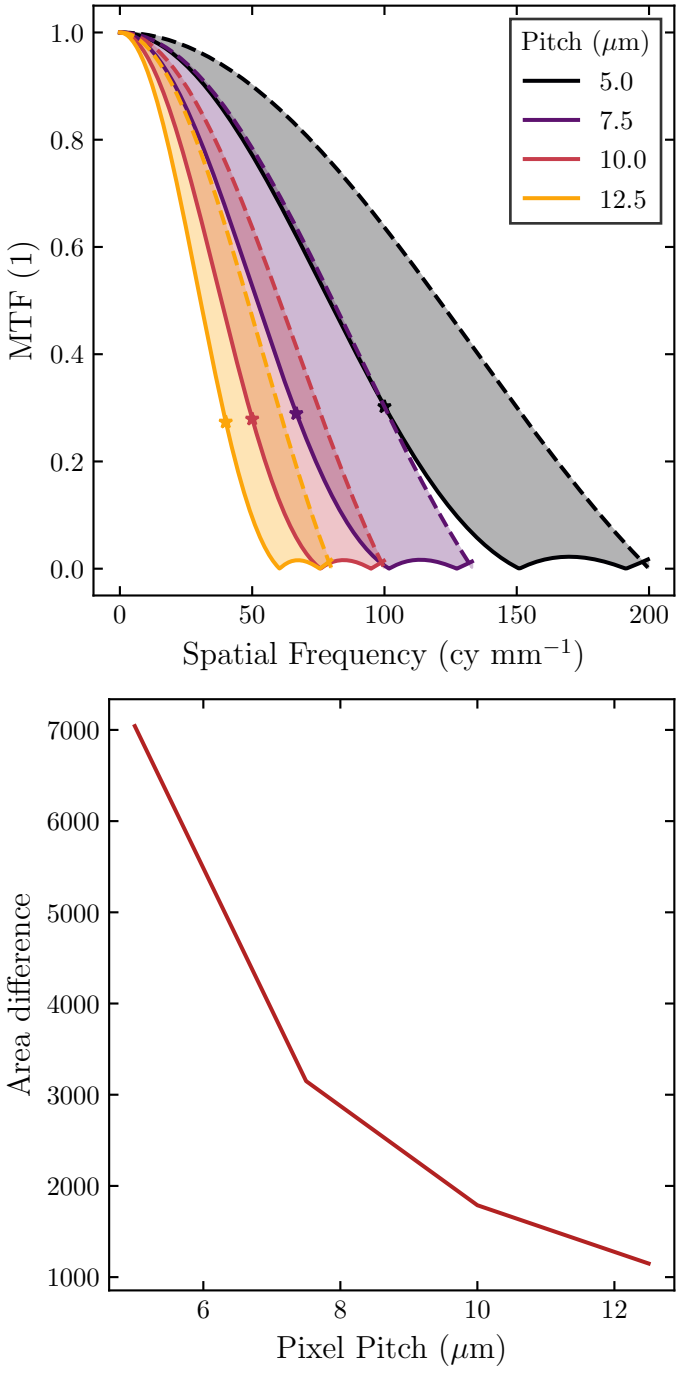

Figure 4·9: (Top) Simulated MTF versus pitch for constant spatial frequency in the y-direction. Dashed lines indicate the detector footprint, solid lines are the simulation results, and the star indicates the predicted MTF at the Nyquist frequency. (Bottom) Total difference in area between the detector footprint and simulation results, taking into consideration both spatial frequency components.

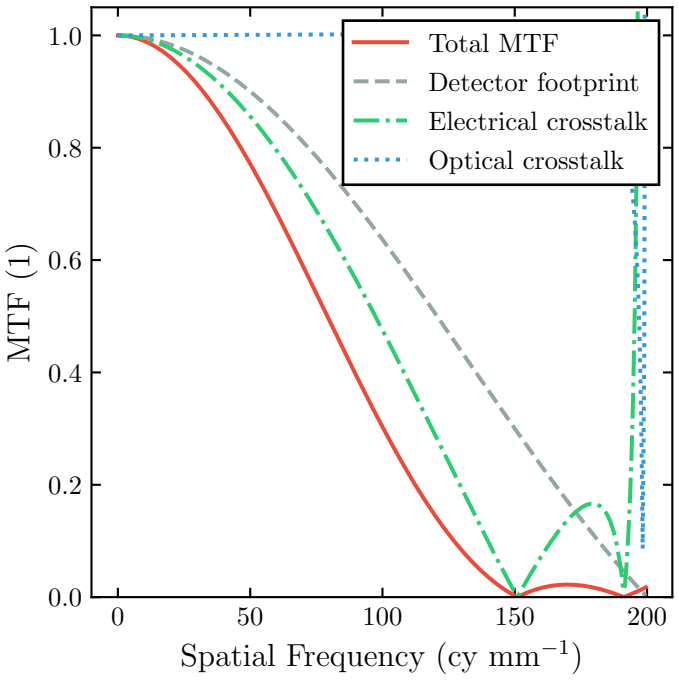

Figure 4·10: Isolated contributions to the MTF.

a 2x2 focal plane array.[4] Points were sampled in a triangular region from $x = 0$ to $x =$ pitch and $y = 0$ to $y =$ pitch, and using the symmetry in the array were reflected and translated to form a full spot-scan of a 3x3 array. The full MTF, shown in Fig. 4·8, was computed using a two-dimensional Fourier transform of this spot-scan profile.

This process was repeated for varying pixel pitch. The results are shown in Fig. 4·9 for constant spatial frequency in the y-direction, or $\eta = 0$. In agreement with the detector footprint improving, the MTF is higher for smaller pixel pitches. The MTF values at the respective Nyquist frequencies are nearly constant. However, while the sensor's MTF improves with smaller pixels, the MTF itself is worse when compared to the expected footprint. To quantify, we subtract the areas beneath the 2D footprint

[4]The calculated MTF from a 2x2 was shown to be in agreement with a representative 3x3 case.

and the 2D simulated MTF,

$$\text{Area difference} = \int_0^{1/\text{pitch}} \int_0^{1/\text{pitch}} (\text{MTF}_{\text{footprint}} - \text{MTF}_{\text{sensor}})\, d\xi d\eta. \tag{4.1}$$

The result is shown in Fig. 4·9; as the pitch is reduced the sensor's MTF is further degraded when compared with the expected footprint.

Using the methodology outlined in Chapter 2, we can isolate the individual contributions to the MTF, as shown in Fig. 4·10. It is observed that optical crosstalk is negligible, and the primary source of MTF degradation is from electrical, or diffusive, crosstalk, as expected for a small-pitch planar array.

4.5 Conclusion

In this chapter we introduced a numerical model for simulating $In_{0.53}Ga_{0.47}As$ focal plane arrays. The model was used to describe how the performance of the sensor is affected by transitioning to smaller pixels. It was shown that the total dark current decreases for smaller pixels, as the lateral contribution to the diffusion current is suppressed due to the interaction between neighboring diodes. While the total dark current decreases, the dark current density increases. As the pixel pitch is reduced, the photodiode is increasingly limited by the vertical diffusion current from below the junction and GR current, which are both pitch invariant. The consequence is that the specific detectivity decreases marginally, by less than a factor of two. The total modulation transfer function of the pixel array improves, but not as much as would be expected when compared to the detector footprint. It was shown that the dominant source of MTF degradation can be explained by inter-pixel diffusive crosstalk.

Chapter 5

Diffusion control junctions

Leveraging the insight gained from the analytical and numerical modeling in the previous chapters, we propose an alternate pixel sub-architecture aimed to reduce dark current and crosstalk in planar photodiode arrays. Specifically, it was noted that $In_{0.53}Ga_{0.47}As$ pixels have a significant contribution to dark current from lateral diffusion from the pixel boundaries. Additionally, with the consistent drive to smaller pitch FPAs, the detector suffers a lower MTF than would be expected by the detector footprint due to increasing inter-pixel crosstalk. To this end, we propose the addition of "diffusion control junctions" (DCJs) into the pixel sub-architecture, with the goal of suppressing the lateral dark current, and mitigating inter-pixel crosstalk. This chapter explores the impact from implementing DCJs into the pixel architecture, and discusses dark current, QE, D^* and MTF.

5.1 Introduction to diffusion control junction concept

Shown in Fig. 5·1 is one implementation of DCJs into the pixel architecture. For this work we are considering a 10 µm pitch, 3 µm thick 10^{16} cm^{-3} n-type absorber, with am abrupt 10^{18} cm^{-3} p^+ cylindrical diffusion that persists 200 nm into the absorber. The main sensing junction has a radius of 2 µm and an applied reverse bias of −200 mV. On the pixel corners are four additional cylindrical abrupt junctions with the same doping density and bias as the sensing junction. The radius of the junctions was chosen to be half of the original sensing junction. Each DCJ is held at the same

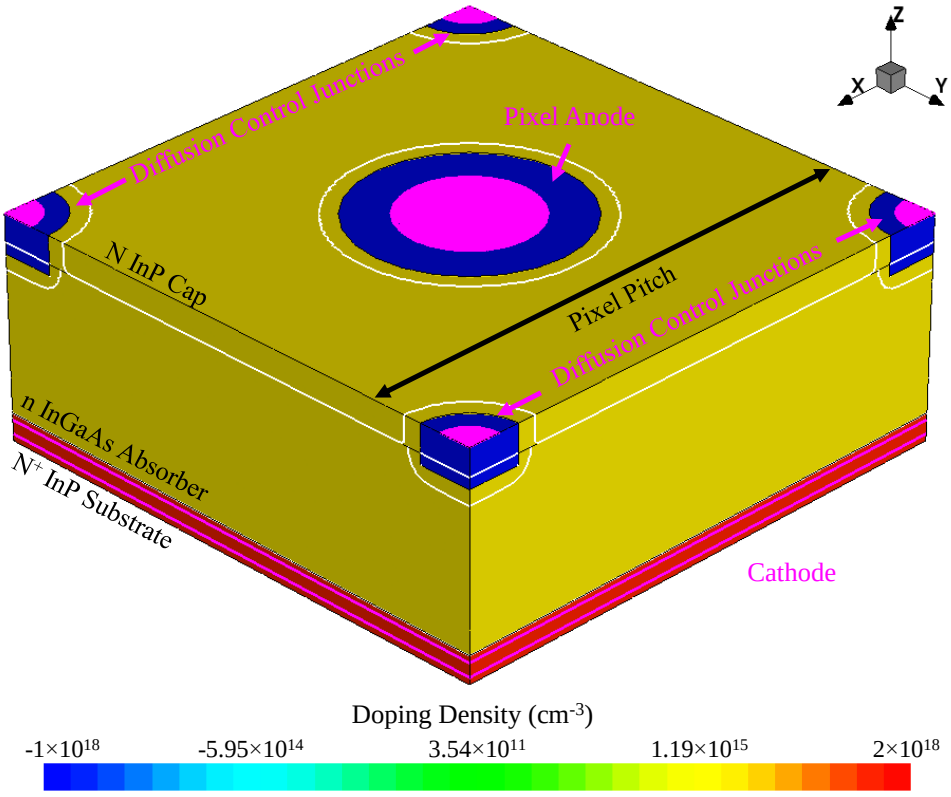

Figure 5·1: $In_{0.53}Ga_{0.47}As$ photodiode with DCJs implemented into the pixel architecture.

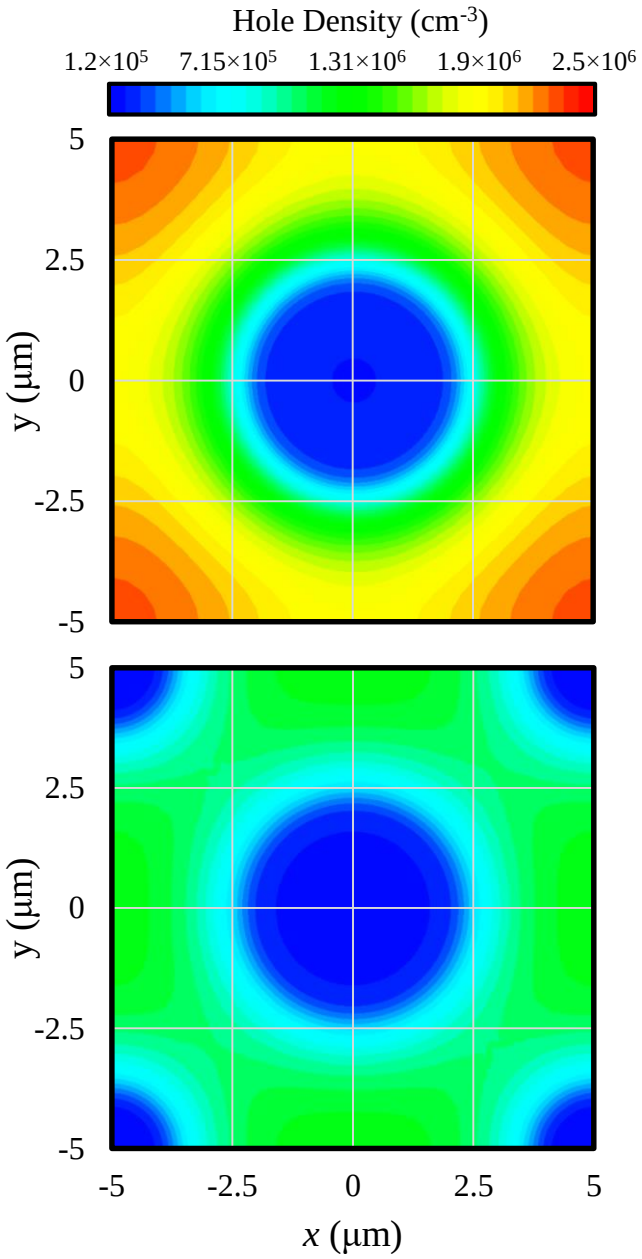

Figure 5·2: Lateral minority carrier density 100 nm below the depletion region edge in the xy-plane. (Top) baseline photodiode; (bottom) photodiode with DCJs located on the pixel corners.

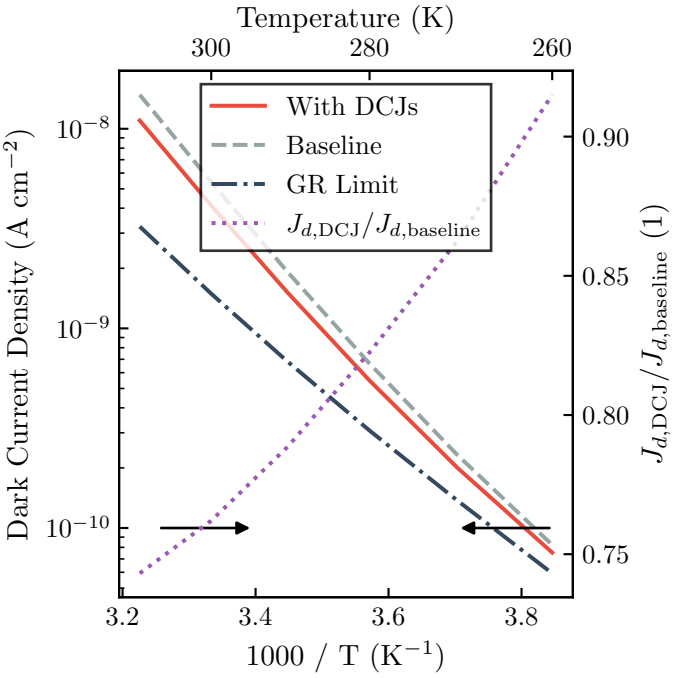

Figure 5·3: Arrhenius plot comparing the dark current density of a baseline photodiode with one with DCJs located on the pixel corners. $V_a = -200$ mV.

reverse bias as the central sensing junction.

The purpose of including additional junctions is to leverage the interaction between neighboring diodes to suppress the minority carrier profile throughout the device. Figure 5·2 shows the minority carrier profile in the xy-plane 100 nm below the depletion region for a baseline photodiode and for a photodiode with DCJs implemented. Observed in the baseline device is a minimum in the minority carrier profile directly beneath the depletion edge, with a rising concentration away from the junction. The corners being the furthest location from the junction have the peak minority carrier concentration. When the DCJs are implemented, there is a substantial suppression of the minority carrier profile laterally throughout the diode. Instead of having peak concentration values at the corners, there are new minima from the DCJ depletion regions, just like what was observed for the p-n-p junctions studied in

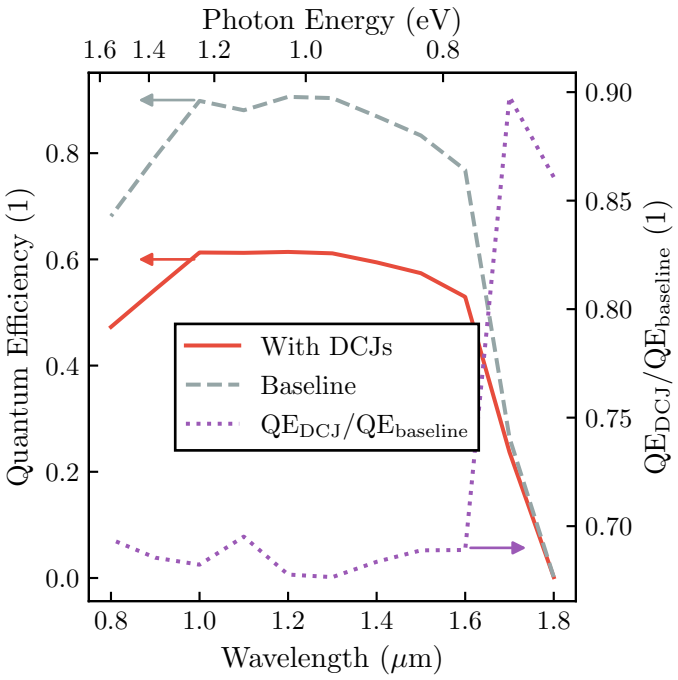

Figure 5·4: Simulated quantum efficiency with and without DCJs and the corresponding ratio with the baseline structure. $T = 300$K, $V_a = V_{\mathrm{DCJ}} = -200$ mV.

Chapter 3. Since the minority carrier current density is proportional to the gradient of the minority carrier concentration, the diffusion current is suppressed as well.

The dark current performance of a photodiode with DCJs implemented is compared with a baseline diode is shown in the Arrhenius plot in Fig. 5·3. For this case, the dark current is reduced by less than a factor of two at high temperatures. At a 10 µm pitch and 2 µm junction radius the lateral contribution to the dark current is comparable to the vertical contribution; a larger pixel pitch would experience a higher degree of dark current reduction as the lateral component is more significant (Glasmann et al., 2017). At low temperatures, the effect is diminished as the photodiode becomes GR-limited; the effect of introducing DCJs into the photodiode is a purely diffusive phenomenon.

It is necessary to also study how the optical performance of the photodiode is

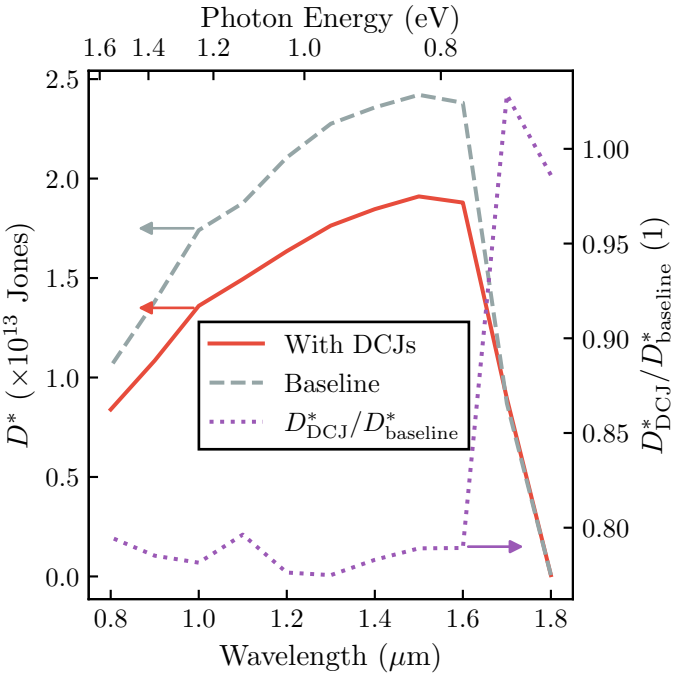

Figure 5·5: Simulated specific detectivity with and without DCJs and the corresponding ratio with the baseline structure. $T = 300$K, $V_a = V_{\mathrm{DCJ}} = -200$ mV.

affected by the DCJs. While the DCJs lower dark current by collecting minority carriers from the pixel boundaries, it can be expected that the photocurrent will also be lowered. This is reflected in Fig. 5·4 where the QE is reduced by about 20% when compared with the baseline device. While the QE is lower, the dark current is also lower, therefore a more relevant figure of merit to gauge the impact of the DCJs on the detector performance is D^*. Shown in Fig. 5·5, D^* is also reduced in the device with DCJs, but only marginally—less than a factor of two.

While one goal of the DCJs was to lower dark current with the aim of improving D^*, the other purpose was to act as an effective mitigation strategy to inter-pixel crosstalk from charge carrier diffusion in planar devices. To this end, we computed the MTF. The two-dimensional spot-scan profile is presented in Fig. 5·6. When compared to the baseline device without DCJs, there is a considerable reduction

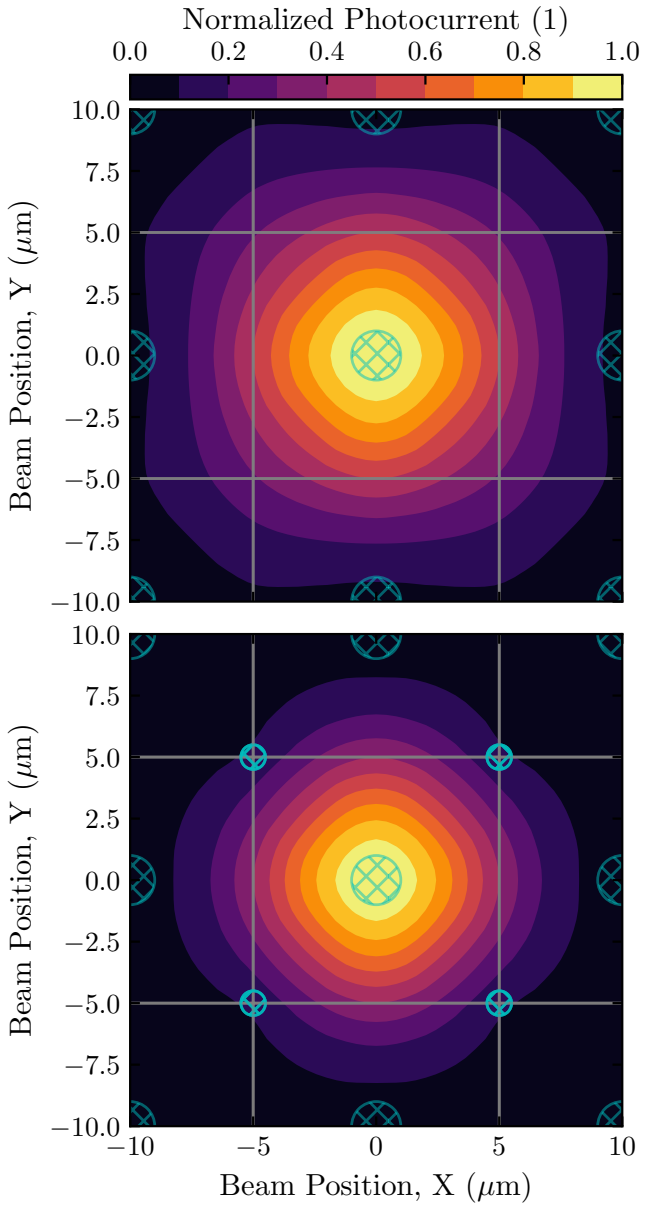

Figure 5·6: Two-dimensional spot-scan profile for a (top) baseline array and a (bottom) array with DCJs included on the pixel corners. The blue hatched circles indicate p^+ diffused regions, and the gray lines denote the pixel boundaries. $T = 300\text{K}$, $\lambda = 1.5$ µm, $V_a = -200$ mV.

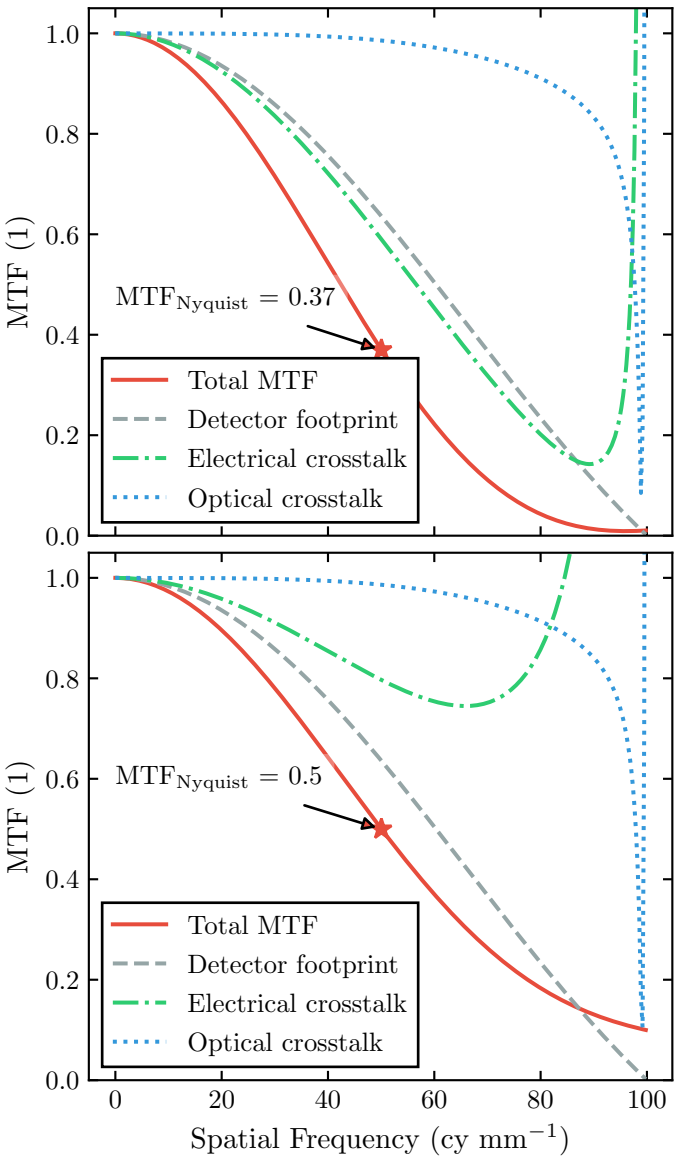

Figure 5·7: MTF and the components for the (top) baseline and (bottom) device with DCJs included on the corners of the pixel. $T =$ 300K, $\lambda = 1.5$ µm, $V_a = -200$ mV.

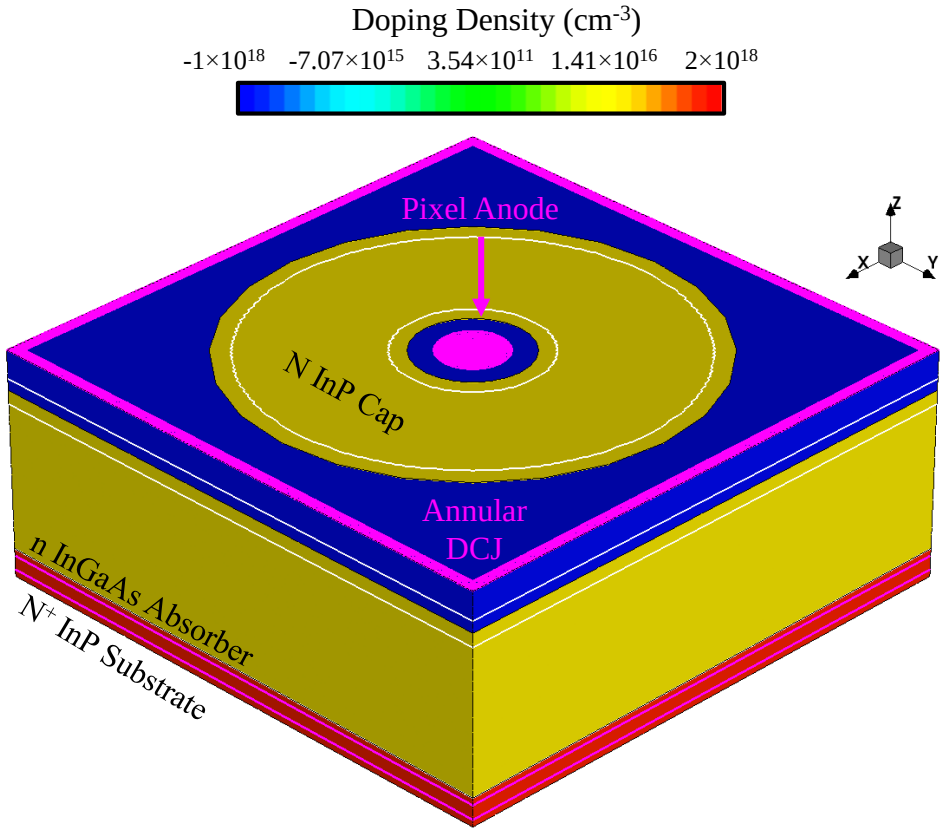

Figure 5·8: An $In_{0.53}Ga_{0.47}As$ pixel with an annular p^+ diffused region as a DCJ.

in photocurrent measured in the central pixel when the Gaussian beam is incident on the neighboring pixels. This is also reflected in the computed MTF in Fig. 5·7. The value of the MTF at the Nyquist frequency is improved from 0.37 to 0.5 when DCJs are present. Decomposing the MTF into the sub-components reveals that the improvement in MTF is due to a reduction in inter-pixel diffusive crosstalk.

5.2 Annular DCJ

In an effort to improve the electro-optical performance we also studied another geometry for the diffusion control junctions. Shown in Fig. 5·8, the DCJs on the pixel corners has been replaced with an annular p^+ diffusion. The intent is to exploit radial

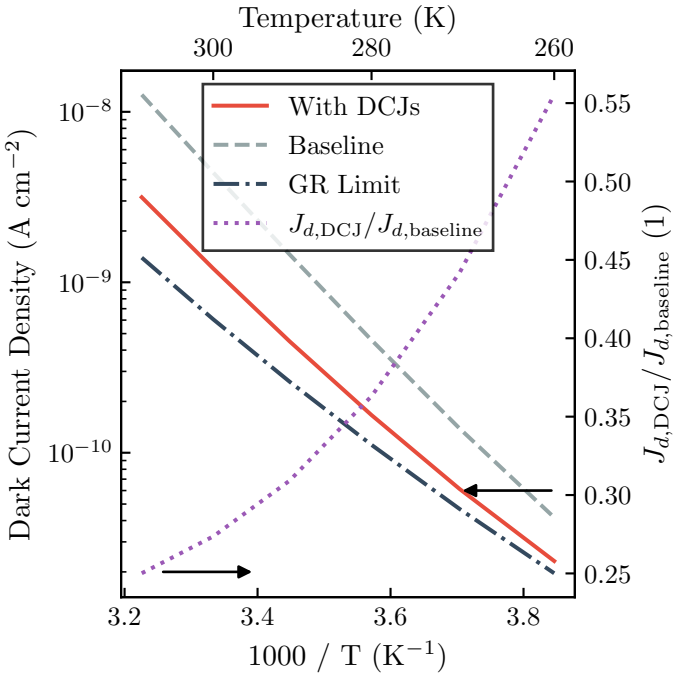

Figure 5·9: Arrhenius plot comparing the annular DCJ geometry in Fig. 5·8 with a baseline device. $V_a = -200$ mV.

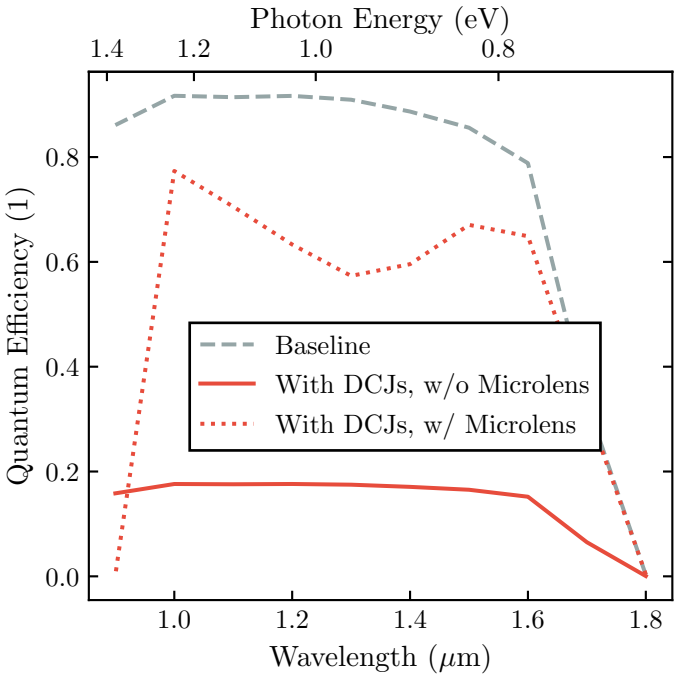

Figure 5·10: Simulated QE for the annular DCJ geometry in Fig. 5·8 with and without a monolithically integrated microlens compared to a baseline device. $T = 300$K and $V_a = -200$ mV.

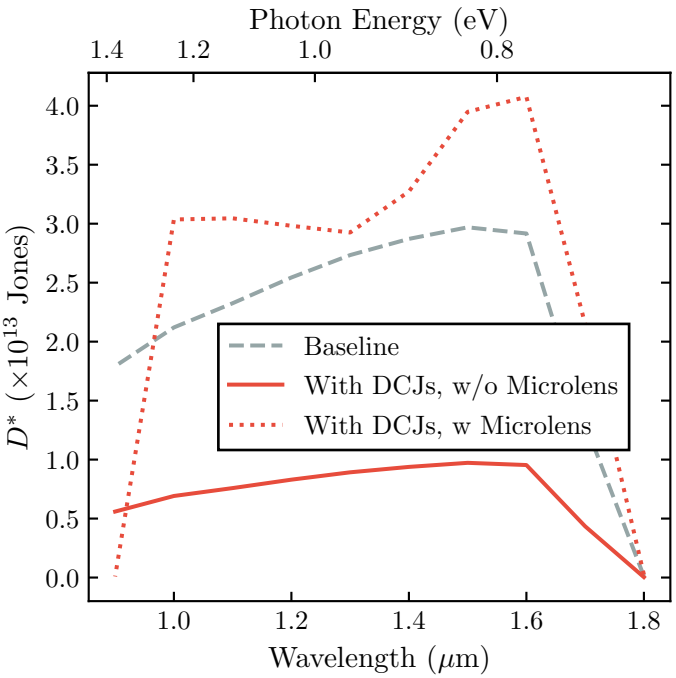

Figure 5·11: Simulated QE for the annular DCJ geometry in Fig. 5·8 with and without a monolithically integrated microlens compared to a baseline device. $T = 300$K and $V_a = -200$ mV.

symmetry to further suppress the minority carrier profile, and lower the dark current further. The structure is the same as the previously discussed device, with the exception that the inner sensing junction has a radius of 1 µm instead of 2 µm. The radius to the edge of the annular is DCJ is 4.5 µm.

Figure 5·9 shows that the new structure is indeed better at lowering the dark current. At high temperatures the dark current in the annular device is close to an order of magnitude lower than the baseline device. Similar to the previous case, the effect is diminished at low temperatures where the GR current begins to dictate the dark current performance. While the dark current is suppressed to a much higher degree, the photocurrent also suffers. In Fig. 5·10 the QE is significantly reduced with values below 20%. The specific detectivity, shown in Fig 5·11, is also more affected than the previous case, with values four times lower than the baseline device.

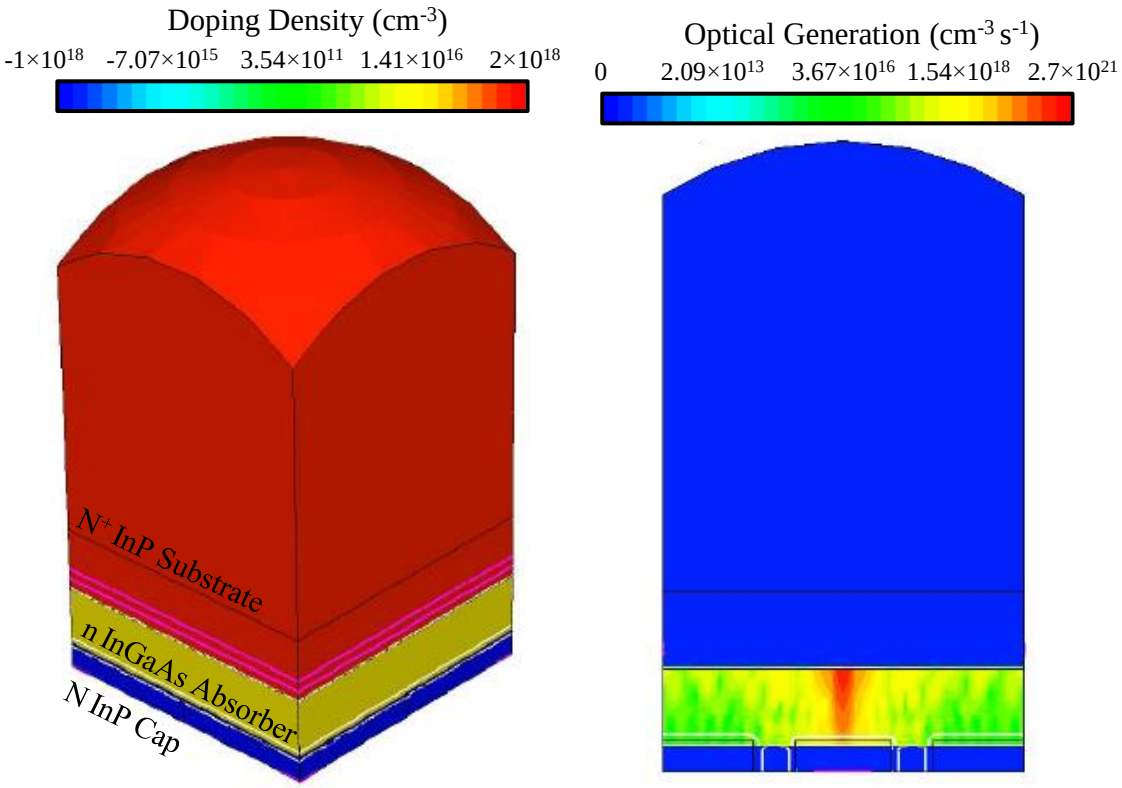

Figure 5·12: (Left) Three-dimensional device model used to simulate the performance of a $In_{0.53}Ga_{0.47}As$ pixel with a microlens fabricated directly on the InP substrate. (Right) Optical generation in the InGaAs absorber vertically through the device. Figure credit: Mr. T. Hubbard.

In an effort to maintain the benefits of lower dark current, and recover the photocurrent, we assessed whether a monolithically integrated microlens would mitigate the QE loss. Shown in Fig. 5·12 is the device structure used in the calculation where the InP substrate is extended, and a spherical microlens is formed. The microlens is shown to focus the radiation towards the center of the pixel, near the original sensing junction as desired The QE performance is improved, as shown in Fig. 5·10, but still lower than the baseline device. However, when converted into D^*, the new annular DCJ geometry with a microlens is better than the baseline device, showing a higher specific detectivity of slightly less than a factor of two. Note, that the monolithically integrated microlens prevents the advantages of a thin substrate, and limits the photoresponse at shorter wavelengths.

5.3 Practical implementation into existing FPAs

Missing until this point is discussion on how the DCJs can be implemented into existing FPA process cycles. The feasibility of implementing additional junctions into the array is diminished if an individual contact to each DCJ is required, therefore we place a constraint that the DCJs must be connected by a common pad, mitigating any complications of adapting the technology to the existing ROIC. We propose two approaches. The first would be to form the DCJs during the zinc diffusion step that forms the original photodiode junctions. If all of the DCJs in the array form a continuous p^+ region, such as in the aforementioned annular structure, then the DCJs can be biases using a common contact located outside of the photodiode array. The second approach would be to use a shallow P^+ diffusion in the InP cap to connect the deeper, discrete DCJs. This would permit the first DCJ strategy to be implemented where they were located on the pixel corners, and allow more strategic placements of DCJs as desired. The first strategy is advantageous from a fabrication

perspective, as no additional process steps are required; only a modification of the masks related to diffusion and p-type contacts would be required. However, the annular structure required a more complex device structure to localize the optical generation to maintain high D^*. The second strategy would require an additional diffusion, adding complexity and masks to the process cycle, but may offer more design flexibility, less QE loss while still improving MTF.

5.4 Conclusion

In this chapter we reviewed a proposed change to the pixel sub-architecture with the goal of reducing dark current and mitigating diffusive crosstalk. We showed that by including additional junctions on the corners of the pixel, the dark current can be reduced. However, the junctions also reduce the photocurrent and result in an appreciably lower QE. However, while D^* was shown to be marginally lower, the MTF of the detector is improved from 0.37 to 0.5 at Nyquist. We also showed that by switching from abrupt cylindrical junctions on the corners of the pixel to an annular p^+ diffused region that the dark current can be lowered by up to an order of magnitude. This approach, however, requires a method of localizing the optical generation near the center of the pixel. We showed that monolithic microlenses fabricated on the back of the InP substrate offers a way of maintaining the benefit of reduced dark current while mitigating the lost QE, enabling a higher D^*.

Chapter 6

Capacitance-voltage characteristics of nBn photodetectors

This chapters presents an extensive analysis of the capacitance-voltage (C-V) characteristics of nBn photodetectors with N- and P- type barrier layers. First, we introduce the device structure and operation. Then, a brief review of C-V profiling of metal-oxide-semiconductor capacitors is given, since the C-V characteristics of barrier-style devices have many similarities. Third, we develop a semi-analytical approach of calculating the capacitance, and apply the model to data available in literature. Lastly, a comprehensive parametric study of the C-V characteristics is presented alongside physics-based discussion of the results. The results and discussion in this chapter have been adapted from (Glasmann et al., 2019).

6.1 Introduction to barrier detector operating principles

Before continuing in discussing the C-V characteristics of barrier-style devices, it is useful to review the basic device structure and operation. Figure 6·1 presents

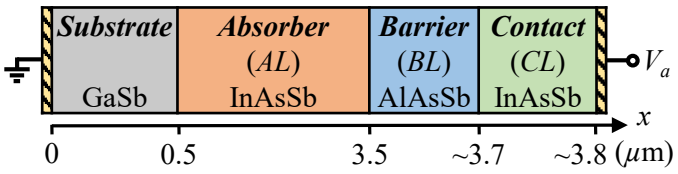

Figure 6·1: One-dimensional nBn structure considered in this work.

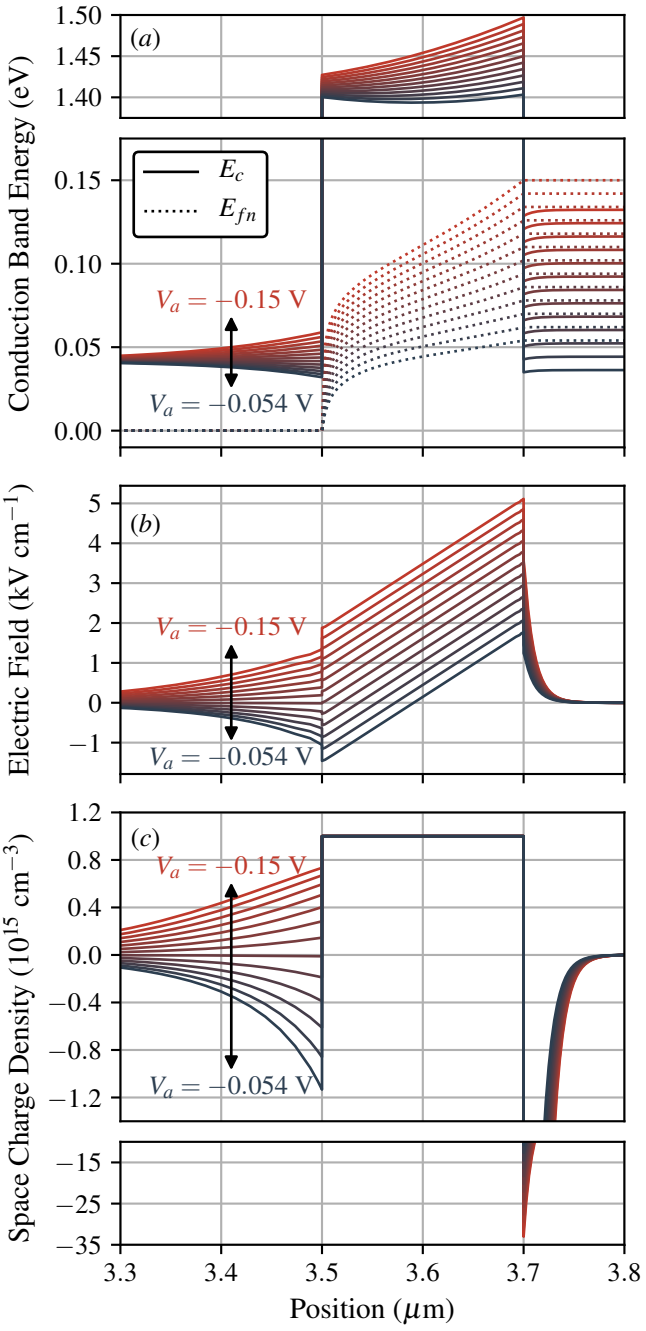

Figure 6·2: Simulated (a) conduction band diagram, (b) electric field, and (c) space charge density plotted as a function of position. The AL-BL interface is at 3.5 µm and the BL-CL interface is at 3.7 µm.

Table 6.1: Device Structural Parameters

Layer	Thickness (µm)	Doping Density (cm^{-3})
Substrate	0.5	Dependent on absorber
Absorber	3.0	$10^{14} - 10^{17}$
Barrier	$0.1 - 0.3$	$10^{15} - 10^{17}$
Contact	0.1	$5 \times 10^{15} - 10^{17}$

the device considered in this chapter. The epitaxial stack consists of an idealized GaSb substrate where the doping density and electron affinity are chosen to ensure zero conduction band offset with the absorber, a 3 µm $InAs_{0.91}Sb_{0.09}$ absorber, an $AlAs_{0.1}Sb_{0.9}$ barrier, and a 100 nm $InAs_{0.91}Sb_{0.09}$ contact layer.

The basic operation of an nBn device with an N-type barrier is shown in Fig. 6·2. At equilibrium the absorber-barrier and barrier-contact interfaces are under slight accumulation. As a reverse bias is applied to the contact layer with respect to the grounded substrate, the absorber-barrier interface moves from accumulation to flat-band and eventually depletion, as shown in the conduction band diagram and space charge density plots. When the polarity of the applied voltage is swapped, the opposite occurs. Instead, the contact-barrier interface moves from accumulation, to flat-band, and depletion, while the absorber-barrier interface is under accumulation.

6.2 Capacitance-voltage profiling

Capacitance-voltage profiling is a common technique for non-destructively characterizing semiconductor devices. Often, C-V profiling is used to measure the doping density in p-n junctions and metal-oxide-semiconductor (MOS) capacitors. More details regarding the experimental technique, and theory, can be found in other texts (Schroder, 2015, Sze and Ng, 2006).

It can be shown that the full expression for the high-frequency junction capac-

itance from the depletion or accumulation of an n-type semiconducting layer in an MOS device is (Casey, 1999, Sze and Ng, 2006)

$$C_j(V_s) = \frac{\epsilon_0 \epsilon_r}{L_D \sqrt{2}} \left| \frac{e^{\beta V_s} - 1}{(-\beta V_s + e^{\beta V_s} - 1)^{1/2}} \right|, \tag{6.1}$$

where L_D is the extrinsic Debeye length given by $(\epsilon_r \epsilon_0 kT/q^2 N_D)^{1/2}$, β is q/kT, and V_s is the electrostatic potential at the semiconductor surface. Notice that for sufficiently large negative surface potentials the common expression for parallel plate depletion capacitance in an asymmetric p-n junction is recovered

$$C_d(V_s) = \frac{\epsilon_0 \epsilon_r}{W_d(V_s)} = \left(\frac{q \epsilon_r \epsilon_0 N_d}{-2 V_s} \right)^{1/2}. \tag{6.2}$$

Equation (6.2) is typically rearranged to give the doping density as a function of the slope of C_d^{-2}

$$N_d = \frac{-2}{q \epsilon_r \epsilon_0 d(1/C_d^2)/dV_s}. \tag{6.3}$$

Hence, by calculating the slope of C^{-2} versus voltage, the doping density of the semiconducting layer can be measured (Schroder, 2015, Muller and Kamins, 2003).

The total capacitance of a metal-insulator junction is a series combination of the insulator's capacitance and the semiconductor junction capacitance,

$$C(V_a) = \left(C_j^{-1} + C_i^{-1} \right)^{-1}, \tag{6.4}$$

V_a denotes the applied voltage and C_i is the parallel plate insulator capacitance,

$$C_i = \frac{\epsilon_r \epsilon_0}{t}, \tag{6.5}$$

t is the thickness of the insulator. Therefore, the key to determining the capacitance is determining the relation between the applied voltage and the semiconductor surface potential. For an ideal MOS capacitor this is straightforward, as there is a

built-in potential that shifts the potential at the semiconductor surface based on the difference between the metal work function and the semiconductor's electron affinity. For a barrier detector, however, it is not as straightforward since there are several semiconducting layers that must be taken into consideration in the analysis.

6.3 Analytical modeling of C-V characteristics of nBn devices

Similar to a MOS capacitor, the capacitance of a barrier detector can be expressed as a series capacitance from the three layers,

$$C = \left(C_{j,\mathrm{AL}}^{-1} + C_{\mathrm{BL}}^{-1} + C_{j,\mathrm{CL}}^{-1}\right)^{-1}, \tag{6.6}$$

where $C_{j,\mathrm{AL}}$ and $C_{j,\mathrm{CL}}$ are the junction capacitances from the AL and CL given by (6.1) and C_{BL} is the parallel plate barrier capacitance given by (6.5) from the barrier layer.[1] Equation (6.6) is exceedingly useful in qualitatively describing the capacitance of this style of device. For example, consider a large reverse bias applied to the CL. As discussed previously, and shown in Fig. 6·2, this causes the contact-barrier interface to be under accumulation, and the depletion region to expand into the absorber. As such, the CL junction capacitance is large and its contribution becomes negligible in (6.6) when compared to the depletion capacitance from the AL and parallel plate capacitance from the BL. Then, (6.6) reduces to the previously discussed case of a MOS capacitor, and we need to determine the surface potential of the absorber-barrier interface as a function of the applied voltage to evaluate (6.1).

To determine the surface potential at the absorber-barrier interface we start by applying Poisson's equation to the semiconductor layer (Sze and Ng, 2006). The

[1] Implicit in this definition of C_{BL} is that the barrier is fully depleted. Otherwise, C_{BL} will have to be replaced with two additional junction capacitances at each of the barrier's interfaces. The fully depleted assumption holds well for a wide range of device parameters, but does not hold for higher barrier doping levels, as seen in the simulation results in later sections.

charge density in a semiconductor expressed in one-dimension is

$$\rho(x) = q\left(N_D^+ - N_A^- + p - n\right),$$

where ρ, N_D^+, N_A^-, p, and n are the total charge density, density of ionized donors, density of ionized acceptors, hole density, and electron density respectively. Considering an n-type semiconductor, the above reduces to[2]

$$\rho(x) = q\left(N_D^+ - n(x)\right). \tag{6.7}$$

Assuming Boltzmann statistics, $n(x)$ is given by

$$n(x) = N_D e^{qV_s(x)/kT}.$$

Poisson's equation, applied to the AL, is

$$\frac{\partial^2 V(x)}{\partial x^2} = \frac{q}{\epsilon_{\mathrm{AL}}}\left(N_{d,\mathrm{AL}} e^{qV(x)/kT} - N_{d,\mathrm{AL}}\right), \tag{6.8}$$

where $V(x)$ is the electrostatic potential. Multiplying both sides by $\partial V/\partial x$ and using the following relations from the chain rule

$$\frac{\partial}{\partial x}\left[\left(\frac{\partial V}{\partial x}\right)^2\right] = 2\frac{\partial V}{\partial x}\frac{\partial^2 V}{\partial x^2}$$
$$\frac{\partial}{\partial x}\left[e^{\beta V(x)}\right] = \beta\frac{\partial V}{\partial x}e^{\beta V(x)}$$

(6.8) becomes

$$\frac{1}{2}\frac{\partial}{\partial x}\left[\left(\frac{\partial V}{\partial x}\right)^2\right] = \frac{qN_d}{\epsilon_{\mathrm{AL}}}\left[kT\frac{\partial}{\partial x}\left(e^{qV(x)/kT}\right) - \frac{\partial}{\partial x}\left(V(x)\right)\right]. \tag{6.9}$$

[2]Note, we have neglected the minority carrier density in this analysis. Unlike an n-MOS capacitor the absorber-barrier and contact-barrier interfaces will not invert; there is not a barrier to holes and they can free move through the structure.

Setting $x = 0$ at the absorber-barrier interface, and using the depletion approximation we can integrate (6.9) from $x = x_d$ to $x = 0$, recalling that the electric field, $\mathcal{E}(x) = -\partial V \partial x$, is zero at $x = x_d$ to obtain the following equation for the electric field at the absorber-barrier interface on the absorber side

$$\mathcal{E}(0^-) = -\mathrm{sgn}(V_{\mathrm{AL}})\sqrt{\frac{2qN_{d,\mathrm{AL}}kT}{\epsilon_{\mathrm{AL}}}\left(e^{qV_{\mathrm{AL}}/kT} - 1 - \frac{qV_{\mathrm{AL}}}{kT}\right)}, \tag{6.10}$$

where V_{AL} is the absorber-barrier surface potential.

Next, we apply Poisson's equation to the barrier layer, following a similar analysis by (Reine et al., 2014). Assuming the BL is fully depleted, Poisson's equation is

$$\frac{\partial^2 V(x)}{\partial x^2} = \frac{-qN_{d,\mathrm{BL}}}{\epsilon_{\mathrm{BL}}} \tag{6.11}$$

Integrating (6.11) yields

$$\frac{\partial V}{\partial x} = \frac{-qN_{d,\mathrm{BL}}x}{\epsilon_{\mathrm{BL}}} + C_1, \tag{6.12}$$

C_1 is an integration constant that can be determined by considering the continuity of electric displacement, $\epsilon_{\mathrm{AL}}\mathcal{E}(0^-) = \epsilon_{\mathrm{BL}}\mathcal{E}(0^+)$, giving $C_1 = -\epsilon_{\mathrm{AL}}\mathcal{E}(0^-)/\epsilon_{\mathrm{BL}}$. Integrating (6.12) again gives

$$V(x) = \frac{-qN_{d,\mathrm{BL}}x^2}{2\epsilon_{\mathrm{BL}}} - \frac{\epsilon_{\mathrm{AL}}\mathcal{E}(0^-)x}{\epsilon_{\mathrm{BL}}} + C_2. \tag{6.13}$$

Since $V(0) = V_{\mathrm{AL}}$, $C_2 = V_{\mathrm{AL}}$. The potential on the other side of the barrier at $x = t_B$, where t_B is the thickness of the BL, assuming negligible voltage dropped across the CL[3] is simply the applied voltage plus any built-in potential due to asymmetric doping between the AL and CL. Evaluating (6.13) at $x = t_B$, where t_B is the thickness of

[3]This assumption holds for the currently considered case for a sufficiently large reverse bias applied to the device such that most of the voltage is dropped across the absorber and barrier. Under forward bias, and for a different barrier polarity, this assumption is no longer valid.

the barrier, and rearranging to solve for $\mathcal{E}(0^-)$ gives the second equation

$$\mathcal{E}(0^-) = -\frac{\epsilon_{\mathrm{BL}}}{\epsilon_{\mathrm{AL}} t_B}\left(V_{bi} + V_a + \frac{qN_{d,\mathrm{BL}}t_B^2}{2\epsilon_{\mathrm{BL}}} - V_{\mathrm{AL}}\right). \tag{6.14}$$

By setting $V_{\mathrm{AL}} = 0$, we can use (6.10) and (6.14) to solve for V_a, and determine the flat-band or crossover voltage where the absorber-barrier interface switches from accumulation to depletion:

$$V_x = -V_{bi} - \frac{qN_{d,\mathrm{BL}}t_B^2}{2\epsilon_{\mathrm{BL}}}. \tag{6.15}$$

Equation (6.15) is the same expression derived by (Reine et al., 2014), except for the inclusion of the built-in potential between the AL and CL. Unfortunately, a closed-form solution for V_{AL} does not exist when equating (6.10) and (6.14) and numerical techniques must be used. Once V_{AL} is known, it can be substituted into the original capacitance equations to determine the C-V characteristics.

This approach has been validated by comparing the C-V characteristics to drift-diffusion simulations, and by fitting data found in literature. Shown in Fig. 6·3, the model closely matches the drift-diffusion results over a wide range of parameters. The slight overestimation of the capacitance near zero bias in Fig. 6·3(b) stems from the exclusion of the contribution from the CL.

Shown in Fig. 6·4, the model was used to fit a set of C-V data from literature (Klipstein et al., 2010). The fitted barrier thickness is about 60 nm thicker than the reported value used in their simulation, which is consistent with their model overestimating the capacitance near zero bias where the barrier capacitance given by (6.5) should be dominant while both the AL and CL are accumulated. Exact values of the barrier doping density are not provided, but the relatively high doping predicted by the model is consistent with a large reverse bias required to deplete the absorber. The fitted absorber doping density is consistent with their report.

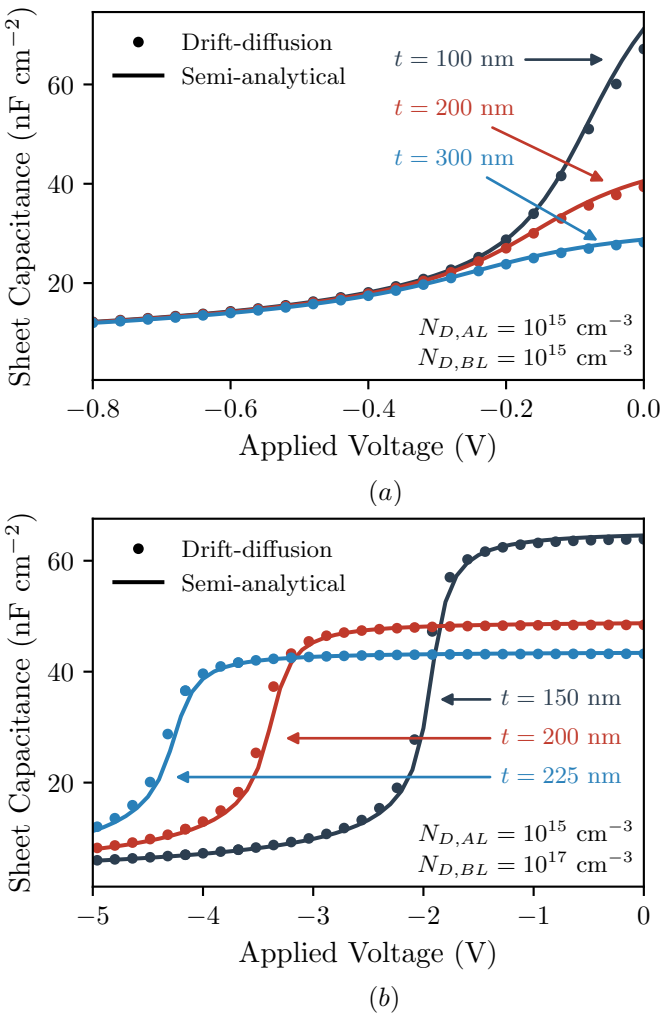

Figure 6·3: A comparison of the semi-analytical approach considered in this chapter to a representative one-dimensional drift-diffusion simulation for varied barrier thickness and an (a) low-moderate doping and (b) high doping density in the barrier layer.

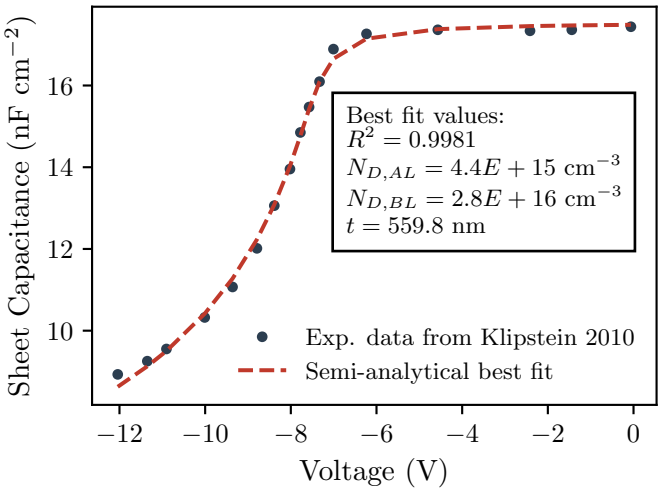

Figure 6·4: A comparison of the semi-analytical approach considered in this chapter to a representative one-dimensional drift-diffusion simulation for varied barrier thickness.

This semi-analytical approach to fitting a C-V profile can be useful for extracting additional information from the C-V characteristics of nBn devices, namely the barrier thickness and doping density. However, the model is not valid for all cases. One limitation is our assumption of a fully-depleted BL. Fortunately, for many case of the N-type BL this assumption holds. However, a high doping density in the BL may cause the assumption of a fully-depleted BL to no longer valid. In these cases, the BL capacitance must be replaced with a series capacitance arising from two disparate depletion regions expanding into a small quasi-neutral region. The manifests as a capacitance that exceeds the parallel plate barrier capacitance "limit." We also neglected the contribution from the CL in determine the device capacitance. For the N-type BL this holds well under reverse bias. Under forward bias the theory in this section could be used to extract the doping density of the CL. However, when considering a P-type BL, this assumption fails since both the absorber-barrier and barrier-contact interfaces begin in depletion, meaning that the CL depletion capacitance is no longer negligible. This is explored in detail in the following section. For

these more complex cases, a complete description of the C-V characteristics could be obtained by applying Poisson's equations to all three layers and solving the resulting set of equations.[4]

6.4 Numerical modeling of C-V characteristics

In this section we review an extensive parametric study to understand the role of the different architectural parameters on the C-V characteristics of nBn photodetectors. The device structure used in the simulations is the same as was shown in Fig. 6·1 with a thickness of 1 µm to represent a one-dimensional device. The simulation frequency and temperature were fixed at 1 MHz and 150K respectively. Unless otherwise mentioned, the contact doping density was fixed at 10^{17} cm^{-3}. The voltage was swept from -8 to 8 V with 101 steps. The capacitance is found using the small-signal sinusoidal steady-state analysis approach implemented by Sentaurus TCAD (Synopsys, 2017a, Laux, 1985).

6.4.1 N-type barrier layer

Figure 6·5 presents several C-V curves for absorber doping densities, $N_{d,\mathrm{AL}}$, between 10^{14} and 10^{17} cm^{-3} and contact doping densities, $N_{d,\mathrm{CL}}$ between 10^{15} and 10^{17} cm^{-3}. Near zero bias the total capacitance is limited by the barrier parallel plate capacitance denoted by the horizontal dotted line, since both the AL and CL are under accumulation. For low $N_{d,\mathrm{AL}}$, the absorber rapidly depletes with increasing reverse bias, lowering the capacitance as $C_{j,\mathrm{AL}}$ dominates in (6.6). The value under strong reverse bias approaches 4.5 nF cm^{-2} consistent with the value predicted by (6.2) for a fully depleted 3 µm layer. The curves with higher AL doping densities will also converge to this value under sufficiently large reverse bias. The same occurs under

[4]Dr. B. Pinkie private communications.

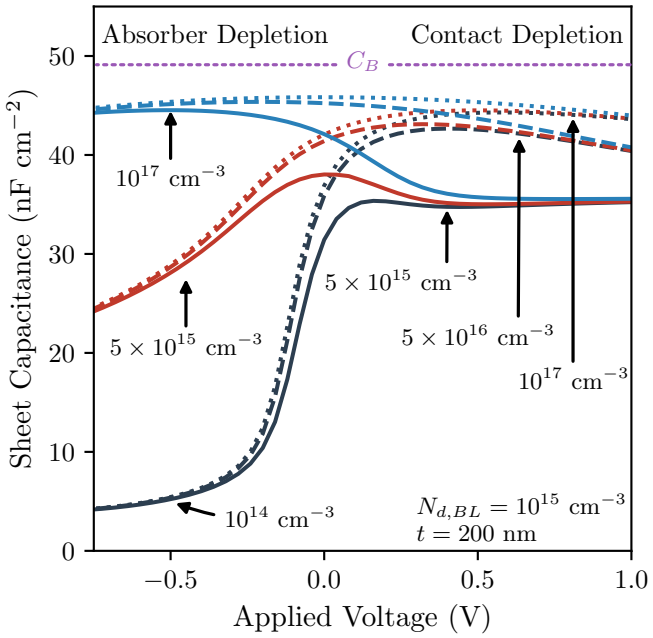

Figure 6·5: Simulated C-V characteristics with an N-type barrier layer for varied absorber and contact doping densities. Lines are styled according to contact layer doping and colored by absorber layer doping.

forward bias where the CL depletes. However, unlike the AL, the CL is quite thin, and fully depletes with low forward bias when $N_{d,\mathrm{CL}}$ is low.

The C-V characteristic as a function of barrier thickness and absorber doping density is shown in Fig. 6·6 for $N_{d,\mathrm{BL}} = 10^{15}$ cm^{-3}. All cases converge to the same large reverse bias value as $C_{j,\mathrm{AL}}$ dominates (6.6). Near zero bias and under forward bias, however, the barrier thickness plays an important role in determining the capacitance value. Since the CL is thin, its depletion capacitance is not low enough to limit the total capacitance, so the parallel plate barrier capacitance that depends on thickness is important. For large values of $N_{d,\mathrm{AL}}$ the barrier thickness has a minor effect. A larger doping density in the absorber causes a larger required reverse bias to deplete, and hence the depletion capacitance is larger for the studied voltage range.

Increasing the barrier doping density affects the C-V profile by increasing the

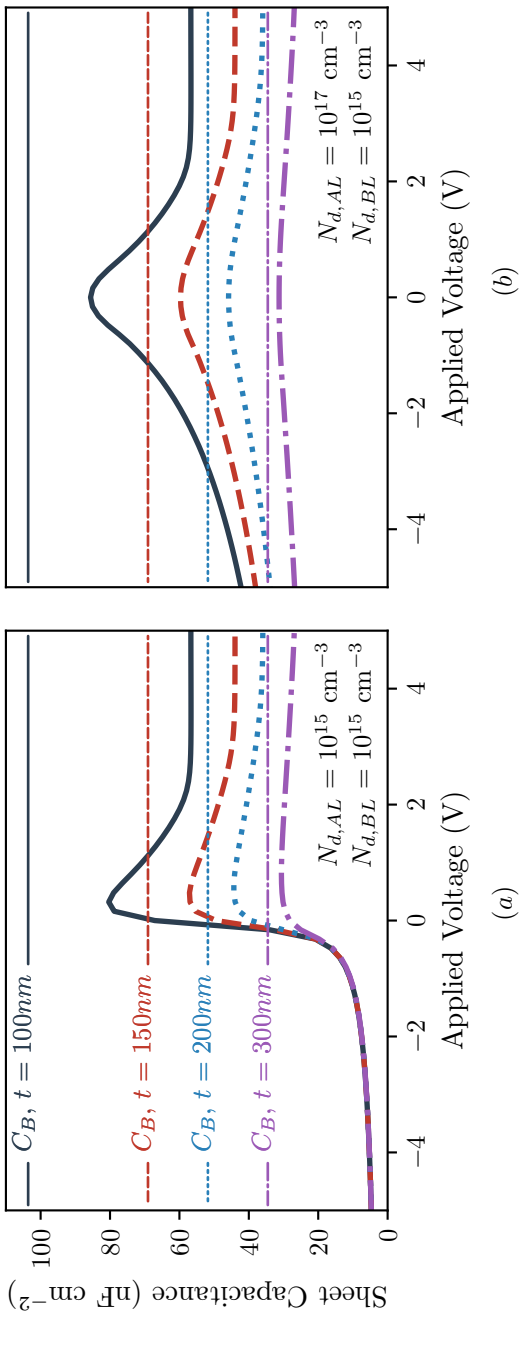

Figure 6·6: Simulated C-V profiles with N-type barrier layers for varied barrier thickness and absorber doping density. $N_{d,\mathrm{BL}} = 10^{15}$ cm^{-3}.

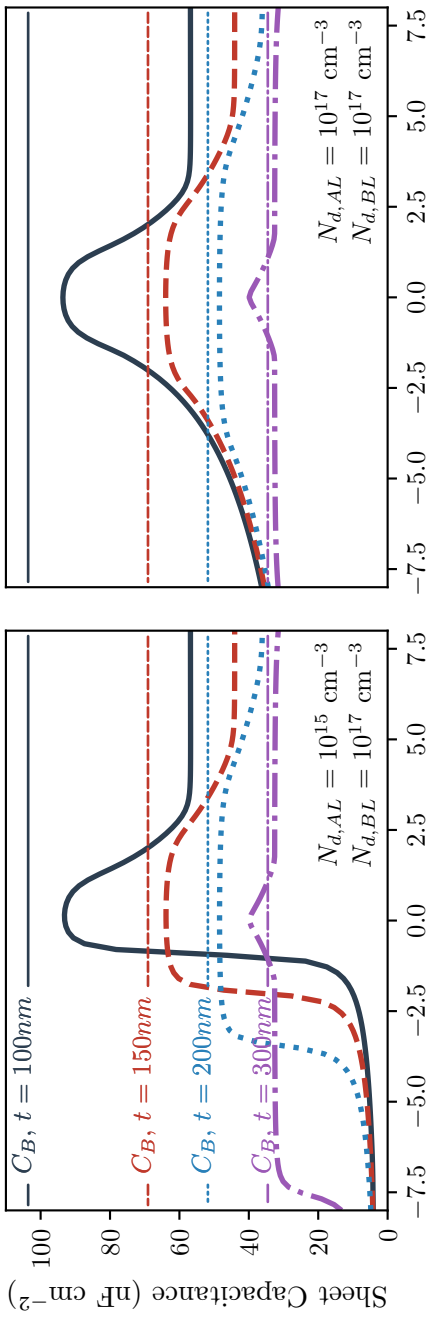

Figure 6·7: Simulated C-V profiles with N-type barrier layers for varied barrier thickness and absorber doping density. $N_{d,\mathrm{BL}} = 10^{17}$ cm^{-3}.

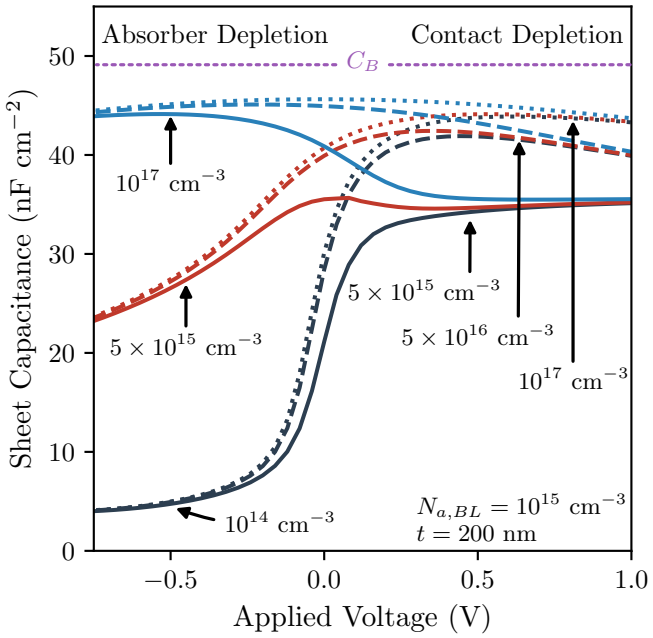

Figure 6·8: Simulated C-V characteristics with a P-type barrier layer for varied absorber and contact doping densities. Lines are styled according to contact layer doping and colored by absorber layer doping.

threshold voltage where the absorber or contact layers switch from accumulation to depletion; the case of $N_{d,\mathrm{BL}} = 10^{17}$ cm^{-3} is shown in fig. 6·7. For the thinner barriers the C-V is broadened at low bias. Unlike the previously discussed cases however, for thick barriers the C-V exceeds the "limit" imposed by (6.5). In these instances, the barrier layer is no longer fully depleted at low bias. Instead, the barrier's contribution to the capacitance is now a series combination of two larger depletion capacitances, which causes the C-V to exceed the parallel plate value. A similar trend is seen in the heavily doped P-type cases.

6.4.2 P-type barrier layer

We now turn to the case of a P-type BL. Figure 6·8 presents the same calculations as Fig. 6·5, except with a P-type BL. The differences between the two figures are subtle, but critical to understand the device operation. A P-type barrier causes the

absorber and contact layers to begin in depletion at equilibrium. This is observed in the curves for $N_{d,\mathrm{AL}} = 10^{14}$ cm^{-3} where there is no longer a maximum, due to the two depletion capacitances from the AL and CL. The similarities of the low doped N- and $P-$ type BL are due to the low crossover voltages, so much of the same behavior at moderate forward and reverse biases holds.

Like the N-type case, we also studied the effect of barrier thickness and doping on the C-V profile for the P-type case. The results for low $N_{a,\mathrm{BL}}$ for $N_{d,\mathrm{AL}} = 10^{15}$ and 10^{17} cm^{-3} are shown in Fig. 6·9. As was the case in the previous discussion of Fig. 6·8, the case of low $N_{a,\mathrm{BL}}$ behaves similarly to low $N_{d,\mathrm{BL}}$. Even though the P-type BL places both the AL and CL in depletion near zero bias, the low crossover voltage means that neither has an appreciable depletion thickness, other than manifesting as a slightly lower capacitance.

When the P-type doping in the BL is increased, the C-V curves become more complex, as presented in Fig. 6·10. The results are explained by considering the case shown in Fig. 6·11. The total capacitance under large reverse bias is limited by depletion in the AL. Near equilibrium at (*i*), there is a local maximum driven by the interplay between a shrinking depletion region in the AL, and the BL capacitance approaching its maximum value from two disparate depletion regions being near equal in size.

The contribution from the BL capacitance decreases as the depletion region at the BL-CL interface expands after (*ii*). Simultaneously, the AL is still depleted and the CL depletes further. These all cause a the local minimum at (*ii*).

Increasing the forward bias from (*ii*) to (*iii*) causes a slight rise in capacitance as the AL depletion shrinks at a faster rate than the depletion region at the BL-CL interface. At (*iii*) is another local minimum; the AL still has a slight depletion layer, but the BL is now fully depleted.

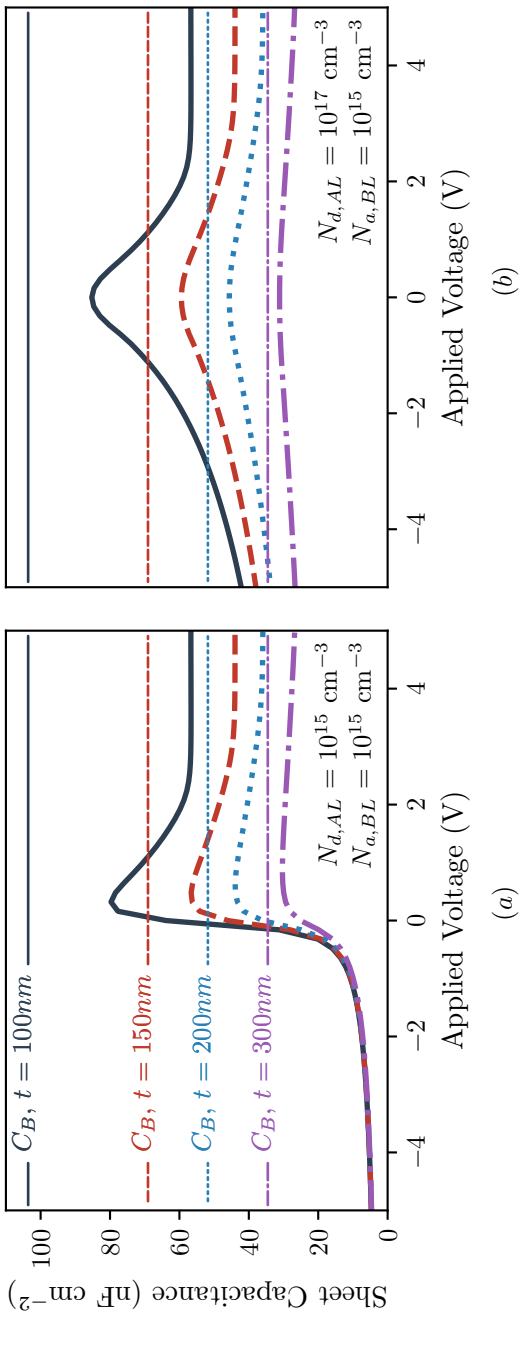

Figure 6·9: Simulated C-V profiles with P-type barrier layers for varied barrier thickness and absorber doping density. $N_{a,\mathrm{BL}} = 10^{15}$ cm^{-3}.

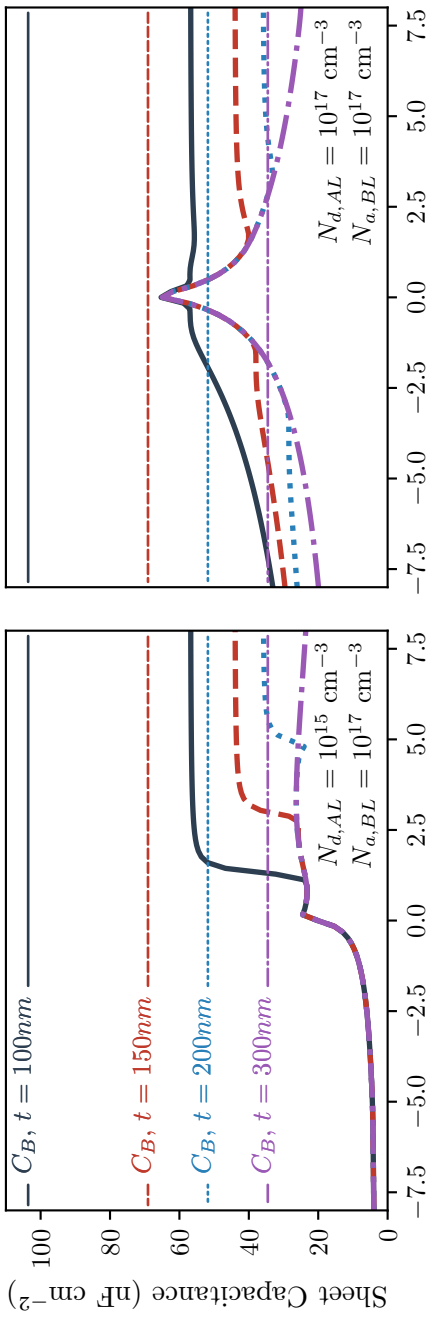

Figure 6·10: Simulated C-V profiles with P-type barrier layers for varied barrier thickness and absorber doping density. $N_{a,\mathrm{BL}} = 10^{17}$ cm^{-3}.

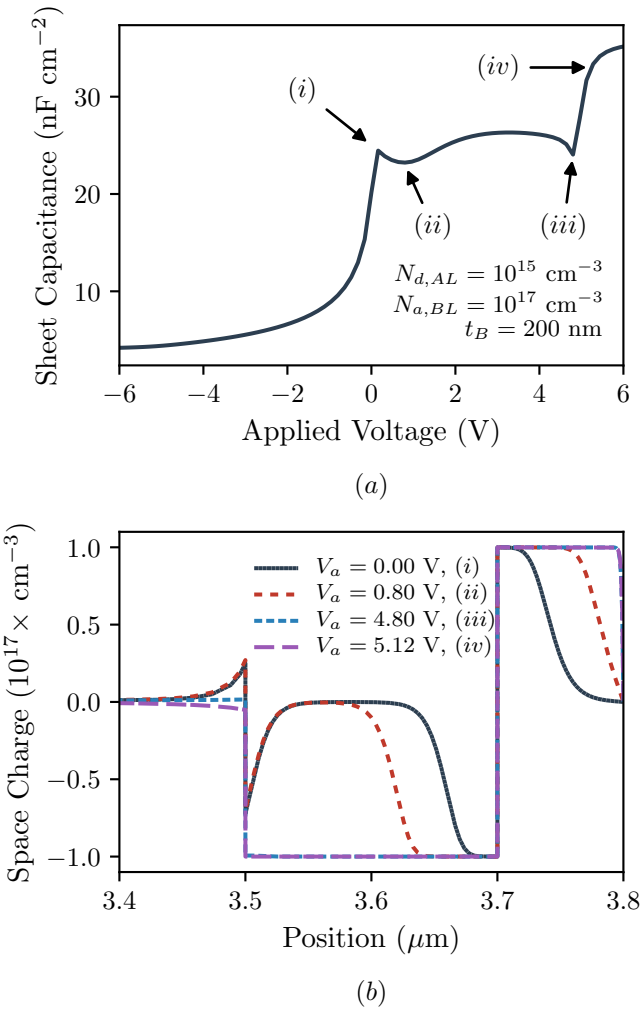

Figure 6·11: (a) Calculated C-V profile for a high P-type doping. (b) Space charge density versus position for the critical voltages noted in (a). The AL-BL interface is located at 3.5 µm and the BL-CL interface is at 3.7 µm.

Finally, the capacitance rises under forward bias towards (iv) as the AL-BL interface accumulates, and no longer contributes to the total capacitance. Notice that the value of the capacitance at 6 V of about 35 nF cm^{-2} is consistent with the forward bias C-V characteristic for the low doped P-type BL in Fig. 6·9.

6.5 Conclusion

This chapter explored the capacitance-voltage characteristics of InAsSb-based nBn detectors with N- and P-type barrier layers. With a foundation built using standard MOS capacitor theory, we showed that the C-V characteristics of nBn devices with N-type barriers can be accurately modeled using a semi-analytical approach. By fitting a C-V profile, it was demonstrated that additional information about the device structure can be obtained, outside of the doping density of the semiconducting layer. The semi-analytical model was validated with a one-dimensional drift-diffusion model, and was applied to experimental data available in literature. Using the drift-diffusion model we performed a comprehensive parametric study to understand the role of each parameter in shaping the C-V characteristic for both N- and P-type barrier layers.

Chapter 7

Neural networks applied to semiconductor device characteristics

Ongoing advances in computing capabilities have led to an unprecedented rise in data availability and creation. Machine learning, a subset of artificial intelligence concerned with studying techniques for building data-driven models, offers a set of tools that are designed to aid researchers in analyzing high-dimensional datasets (Goodfellow et al., 2016, LeCun et al., 2015, Jordan and Mitchell, 2015). Machine learning has been adopted, or has begun to be adopted, for use in many areas of the semiconductor industry. For example, artificial neural networks (ANNs) have long been used to create compact models for circuit simulators (Litovski et al., 1992, Root, 2012, Huang et al., 2016). The goal of incorporating an ANN compact model is to capture a more physically accurate representation of the device; often, the realistic temperature, voltage, or frequency dependence of a device cannot be expressed in closed-form analytic expressions, or they may not be valid under all possible operating conditions or fail to capture all aspects of the device. ANN can offer superior generalization over other approaches, such as lookup tables, and can be used to represent a statistical distribution of fabricated devices subject to different operating conditions. Moreover, once trained, the ANN can be faster to evaluate at runtime when compared to a full-scale numerical approach.

Another compelling application of machine learning techniques is in semiconductor manufacturing. Semiconductor fabrication facilities generate vast amounts of

data that can be leveraged to create data-driven models for fault detection, predictive maintenance, debugging yield issues, and develop enhanced metrology techniques (Irani et al., 1993, Moyne and Iskandar, 2017, Susto et al., 2012, Susto et al., 2015). For instance, recent advances in models for image recognition can be applied to create a system that monitors the progress of a wafer through a process line. Then, through pattern recognition, discover which process introduces defects into the wafer, reducing downtime and revealing a source of low yield. These models can achieve accuracies that can exceed those obtained by human technicians, and may offer a deeper insight into the malfunctioning step of the process.

This chapter is concerned with yet another aspect of the semiconductor industry: applying machine learning to device research and development. A widely encountered challenge in all engineering fields is understanding how each parameter plays a role in determining device performance. Even small-scale problems with only a few design parameters can escape the realm of feasibility when trying to understand parameter-performance correlations, though there have been recent publications of using machine learning techniques to address this problem (Melati et al., 2019). First, we offer a high-level discussion of neural networks. Then, we present two applications of ANNs to the C-V characteristics of nBn photodetectors discussed in the previous chapter. Finally, using ANNs as a surrogate model we explore the tradeoffs of dark current and quantum efficiency for a variety of design parameters, and use the network to assess the tolerance of a design to parameter uncertainties.

7.1 Artificial neural networks

This section is intended to introduce neural networks and the relevant terminology at a high-level; readers interested in learning more about the implementation, algorithms, and other types of networks outside of what is discussed in this work are encouraged

to look elsewhere (Goodfellow et al., 2016, Nielsen, 2015).

There are three main classes of machine learning algorithms: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning trains a model to reproduce an expected result by exposing it to sets of curated, labeled data. Classification and regression are addressed by supervised learning, the former categorizes an input into a set of defined labels, and the latter fits the data over a range of values. Unsupervised learning is used to organize a set of data to highlight trends in large, multi-dimensional datasets, perhaps by clustering data that share common trends. Reinforcement learning tests a model on a certain task, and implements a reward or penalty based on how well it performs. In this chapter, we focus on supervised learning using ANNs for regression.

Artificial neural networks are biologically inspired mathematical models meant to mimic how the brain processes information. The fundamental building block of an artificial neural network is a neuron. The neuron accepts an input, performs a small calculation, and passes the output to subsequent neurons for further processing. A network is constructed by connecting the neurons, where the complete model accepts an input, such as an image, text, or simply raw numeric data, propagates the information through the network where it is manipulated in such a way that the desired output is obtained.

7.1.1 The multi-layer perceptron model

The simplest neural network is based on the multi-layer perceptron model. Shown in Fig. 7·1 is the standard feedforward neural network considered in this work. The network is comprised of three basic layers: an input layer, a hidden layer, and an output layer. The input layer accepts a vector, $\boldsymbol{x} = \{x_1, ..., x_p\}$; the elements of $\boldsymbol{x}$ are often referred to as features or predictors. A series of calculations take place in the hidden layer as information propagates through the network until the final output

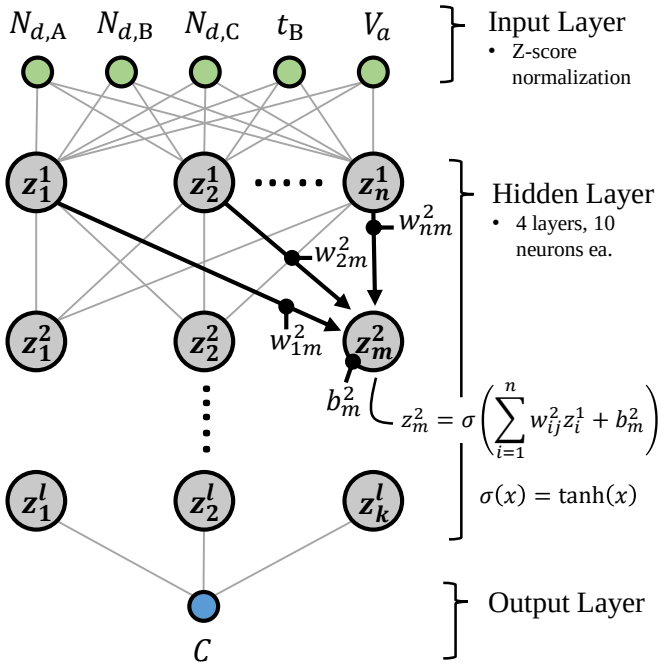

Figure 7·1: The fully-connected neural network considered in this work for predicting the capacitance based on $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, t_{B}, and V_a.

vector, $\hat{\boldsymbol{y}} = \{y_1, ..., y_k\}$, is obtained. The hidden layer is comprised of sub-layers of fully-connected neurons.[1] Consider two sequential layers with n and m neurons respectively. The value given by the j-th neuron in the second layer is the output of an activation function given the weighted sum of the outputs from the previous layer, plus a bias, or

$$z_j^2 = \sigma\left(\sum_{i=1}^{n} w_{ij}^2 z_j^1 + b_j^2\right), \tag{7.1}$$

where w_{ij}^2, b_j^2, and z_j^1 denote the weight associated with the connection between the i-th neuron in the previous layer and the j-th neuron in the current layer, the j-th neuron's bias, and the activated value from the i-th neuron. The function, σ, is the

[1]The architecture of the hidden layers—the number of neurons per layer and number of layers—is often tuned for each application. As a general rule-of-thumb the number of free parameters associated with the network should be similar to the amount of training data to limit the degree of overfitting. Otherwise regularization techniques must be used.

activation function that introduces non-linearity into the model.[2] In this work we will use two different activation functions. The first is the hyperbolic tangent function,

$$\sigma(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}.$$

The second is the exponential linear unit (ELU) given by (Clevert et al., 2015),

$$\sigma(z) = \begin{cases} z & z > 0 \\ e^z - 1 & z \leq 0. \end{cases}$$

As the network complexity is increased by adding more layers and neurons, the output becomes an increasingly complex function of nested function evaluations.

Training the neural network on a regression task is the same as curve fitting, except instead of a simple linear or polynomial function with a few free parameters, a neural network can have hundreds of thousands of unknown variables. The weights and biases in (7.1) for each neuron form the sets of free parameters contained within the model of the neural network. One of the biggest challenges historically in deep learning has been to develop numerically stable and efficient algorithms for optimizing the large number of weights and biases. During the optimization process, usually referred to as training the model, the weights and biases are tuned in such a way that the error in the network's inference when exposed to a subset of the available data is reduced. During training, the network is given N samples, $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}^T$ with a known output, $\boldsymbol{Y} = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_N\}^T$. The output is an array of predictions, $\hat{\boldsymbol{Y}} = \{\hat{\boldsymbol{y}}_1, ..., \hat{\boldsymbol{y}}_N\}^T$. The error is expressed as a cost function representing the total error in the model's inference. Similar to the neurons' activation functions, there are many choices of cost functions that vary based on the problem.[3] Here, we use the

[2]There are many choices of activation functions, and the correct one for the given application is often chosen empirically. Research of new activation functions aimed to lower network prediction error or improve optimization rates is an active area of research.

[3]While not considered here, there has been recent progress in developing cost functions with em-

average value of a loss function, $\mathcal{L}$, for each training sample, or,

$$J(\boldsymbol{X};\theta) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}(\boldsymbol{y_i},\hat{\boldsymbol{y}}_i(\boldsymbol{X};\theta)),$$

where θ is used to denote all hyperparameters used when creating the model, including the weights, biases, learning rates, or even the steps taken during data processing. A common choice of loss function for the regression tasks, like what is considered in this chapter, is to calculate the mean squared error of the prediction given by

$$\mathcal{L}(\boldsymbol{y},\hat{\boldsymbol{y}}) = \frac{1}{k}\sum_{j=1}^{k}(y_j - \hat{y}_j)^2.$$

As mentioned previously, one of the reasons for the recent interest in deep neural networks is due to the emergence of new optimization algorithms for calculating the weights and biases in the network. At the heart of many optimization algorithms is the method of steepest descent, or gradient descent. The algorithm uses automatic differentiation to efficiently find the gradient of J with respect to the weights and biases (Neidinger, 2010, LeCun et al., 2015, Nielsen, 2015). Then, through repeated application of the chain rule, the weights and biases are updated using the gradients such that J ideally reduces with subsequent training iterations; this technique is called backpropagation.

When training the model, the available data is split into training and validation sets. The training data is used to train the model. The validation data, is used to test the generalization ability of the model to unseen samples. Ideally, the error when predicting the validation data should be comparable to the training error. If the validation error is much higher and increases with further training, this usually indicates that the complexity[4] of the model is too high, or there are too many free

bedded physics, such as a conservation law or other relevant equations suited for the given problem. This strategy falls under a rapidly evolving area referred to as physics-informed machine learning.

[4]Here, complexity refers to the number of free parameters in the network, Complexity can be

parameters, and the model is overfitting. Otherwise, if the validation error is high and remains constant with further training, perhaps the complexity of the model is too low, and there is a irrecoverable error between the training and validation. In this case, the model's complexity should be increased, or more data should be gathered.

7.2 C-V modeling with neural networks

In the previous chapter a thorough analysis of the C-V characteristics of InAsSb-based nBn photodetectors was presented, with extensive discussion on the role of the doping densities of the absorber, $N_{d,\mathrm{AL}}$, barrier, $N_{d,\mathrm{BL}}$, and contact, $N_{d,\mathrm{CL}}$ layers, as well as the thickness of the barrier, t_{B}. In this chapter we demonstrate that neural networks can be used as an useful way to explore the multi-dimensional parameter space, without an excessive number of calculations.

One prominent issue when training neural networks, and most numerical modeling problems, is adequate sampling of the feature-space. An easy, but computationally expensive approach, is a grid search, where every permutation of a set of discrete values for each feature is sampled. While simple to implement, and useful for exploring specific combinations of features, the number of unique values shown to the network during training is limited. Furthermore, if we consider using N values for n features, the number of samples required is N^n, which quickly becomes infeasible for large values of N to ensure adequate representation of the parameter space. To address this concern, quasi-random sampling techniques are often used when sampling multi-dimensional spaces. The advantage of switching to a quasi-random sampling technique is better coverage of the parameter space and preventing the possibility of clustering when using random sampling. For the same number of samples in the gridded search, N^n, using a quasi-random sampling approach yields N^n unique values

increased/decreased by adding/removing more layers or neurons.

of each feature, instead of only N. In this work, we have opted for a quasi-random approach based on Halton sequences (Halton, 1960, Halton, 1964, Kocis and Whiten, 1997). The same simulation approach detailed in the previous chapter was used to generate the capacitance-voltage data. Halton sequences were used to discretize $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B} while V_a was fixed at 101 evenly spaced voltages between -8 and 8.

The neural network architecture is shown in Fig. 7·1. The input layer contains five neurons, where values of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, t_{B}, and the applied voltage, V_a are fed to the network. The output of the network is a single neuron with a linear activation representing the inferred sheet capacitance. The hidden layer contains four layers with ten neurons each with hyperbolic tangent activation functions. Weights were initialized using Glorot initialization (Glorot and Bengio, 2010). The training process utilized the Adam optimization algorithm with the recommended values of β_1, β_2, and ε of 0.9, 0.999, and 10^{-7} respectively (Kingma and Ba, 2014).[5] Prior to training, a Z-score normalization was applied to the training inputs to normalize the dataset to have zero mean and unity standard deviation, and the base-10 logarithmic values of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, and $N_{d,\mathrm{CL}}$ were used. For stability over many training epochs the initial learning rate was set to 10^{-4}. The neural network was trained used 128 quasi-randomly sampled values of $N_{d,\mathrm{AL}}$ within 10^{14}–10^{17} cm^{-3}, $N_{d,\mathrm{BL}}$ within 10^{15}–10^{17} cm^{-3}, $N_{d,\mathrm{CL}}$ within 10^{15}–10^{17} cm^{-3}, and t_{B} within 100–300 nm. To improve the model quality at the periphery of the 4-dimensional hypercube of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B}, we also included the 16 corner values of the limiting cases. For example, $N_{d,\mathrm{AL}} = 10^{14}$ cm^{-3}, $N_{d,\mathrm{BL}} = 10^{15}$ cm^{-3}, $N_{d,\mathrm{CL}} = 10^{17}$ cm^{-3}, and $t_{\mathrm{B}} = 100$ nm. This gives a total of 144 simulated designs[6] with 14,544 individual capacitance values over bias. The dataset was split in half to form the training and validation sets by

[5]These values were cross-validated and verified to be reasonable for this problem.

[6]Design refers to a specific combination of architectural properties. In this case, a design has certain values of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B}.

taking every other C-V point as training and every other adjacent point as validation. Since the validation set would always contain the same values of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, t_{B} as used during training, we also created a separate test set with unique values that were not used during training to better assess generalization. Finally, we used TensorFlow[7] as the framework for creating, initializing, training, and evaluating the neural networks.

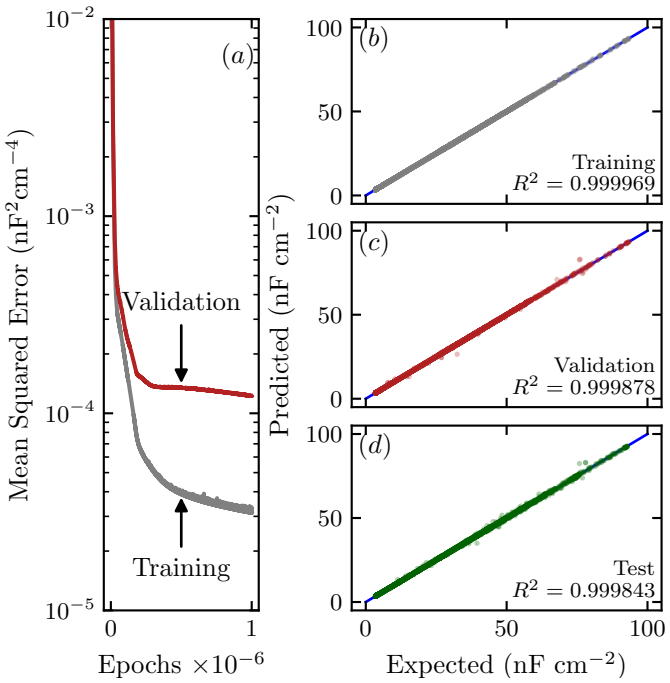

Figure 7·2: (a) Learning curve showing the mean squared error of the capacitance during training. Predicted versus expected capacitance for the (a) training, (b) validation, and (c) test sets. The blue lines have unity slope and zero intercept for reference.

The network was trained over one million epochs; the learning curves are shown in Fig. 7·2. The error in both the training and validation predictions is low without overfitting, indicating a high quality of fit. This is also reflected in the prediction-expectation curves where R^2 values near unity are achieved in all three data sets. This confirms the network generalization ability to unseen cases of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and

[7] `https://www.tensorflow.org/`

t_{B}.

Figure 7·3 also demonstrates the excellent generalization exhibited by the network by looking at univariate parameter sweeps. The network correctly captures the C-V behavior discussed in the previous chapter, and follows the data contained within the test set. In Fig. 7·3(a), the network reproduces the expected behavior that a larger reverse bias is required to deplete the absorbing layer as the doping density increases. In Fig. 7·3(b), the network reproduces the C-V broadening imposed by increasing the doping density in the barrier, causing the crossover voltage where the absorber-barrier and contact-barrier interfaces switch from accumulation to depletion. In Fig. 7·3(c), the network predicts the trends in $N_{d,\mathrm{CL}}$, where larger forward biases are required to deplete the contact layer, and since the layer is thin will eventually fully deplete resulting in a constant capacitance. Finally, in Fig. 7·3(d), the network demonstrates the dominance of the parallel plate capacitance due to the barrier for increasing thicknesses.

Once the network is trained, and verified to yield accurate results and generalize well to other cases, it can be applied as a statistical model to gain a global understanding of the role each feature. One common technique used when analyzing blackbox functions, especially ones that have a larger number of predictors, is to create partial dependence plots (Hastie et al., 2005, Greenwell, 2017, Friedman, 2001). The bivariate partial dependence plots of the capacitance is shown in Fig. 7·4. The plots were generated using the following procedure. Consider Fig. 7·4(a), where the average capacitance is shown as a function of $N_{d,\mathrm{AL}}$ and V_a. First, we fix the value of $N_{d,\mathrm{AL}}$ and V_a. Then, the average capacitance is calculated from the set of predicted capacitances for every permutation of discrete values $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B}, within their respective limits. This value indicates the partial dependence of the capacitance, $\overline{C}$, with respect to $N_{d,\mathrm{AL}}$, by averaging the influence of the other features on the capacitance. This

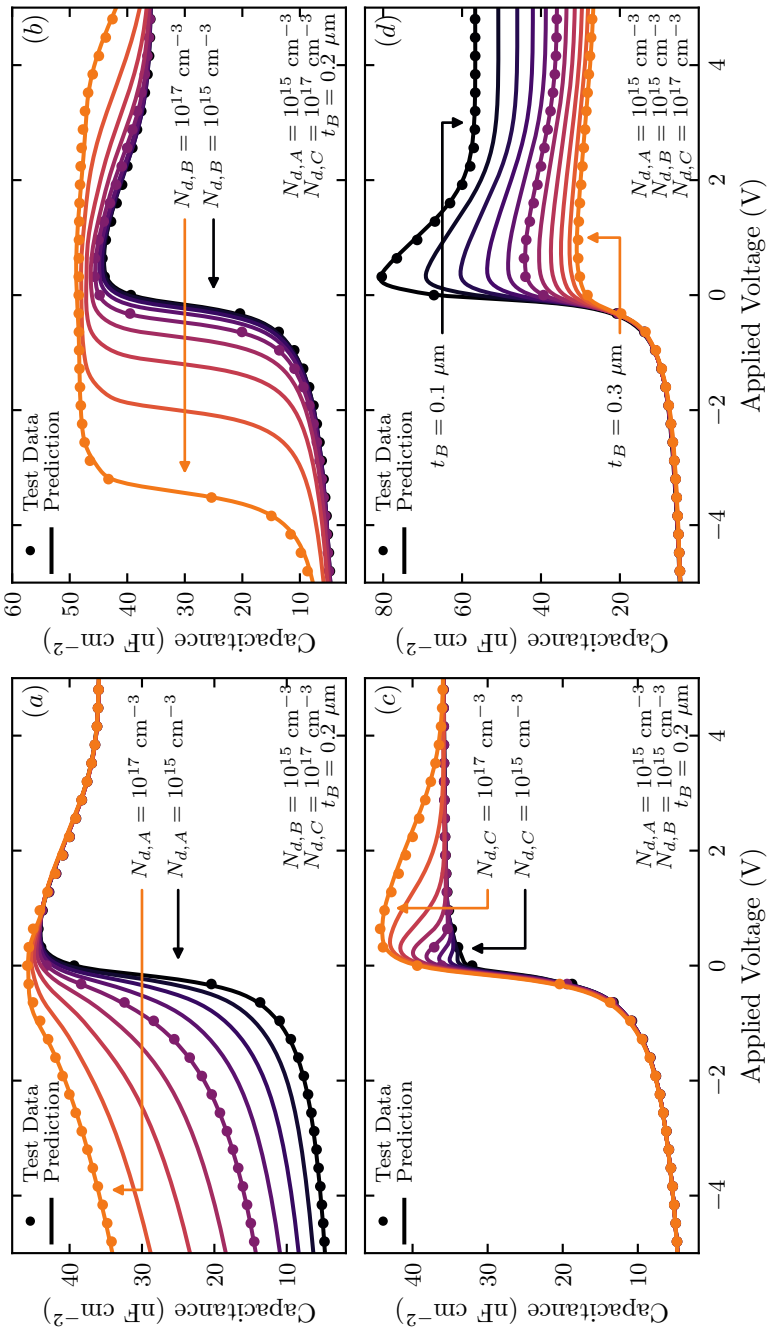

Figure 7·3: Univariate sweeps of (a) $N_{d,\mathrm{AL}}$, (b) $N_{d,\mathrm{BL}}$, (c) $N_{d,\mathrm{CL}}$, and (d) t_{B}. All lines are neural network predictions. Markers indicate data contained in the test set, not used during training.

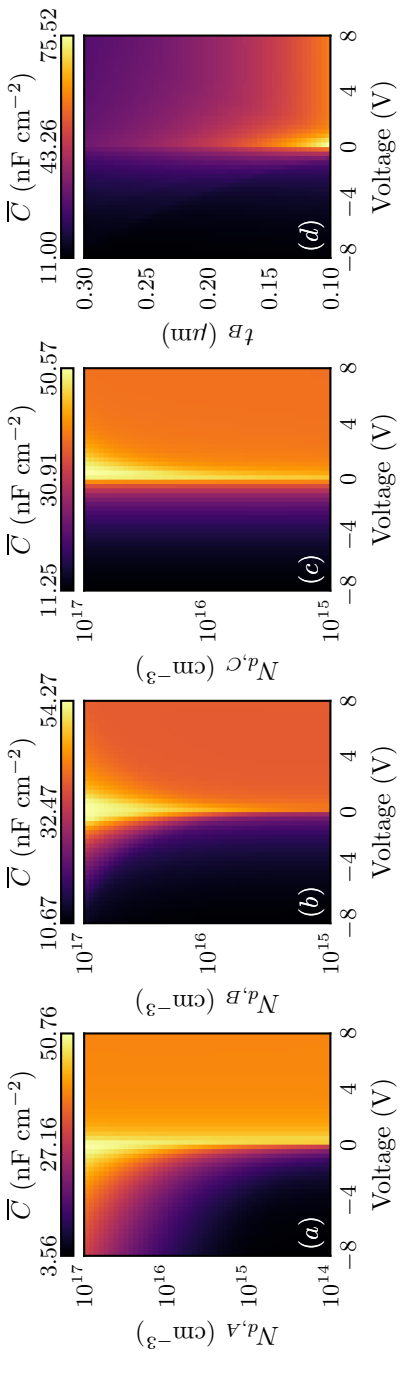

Figure 7·4: Bivariate partial dependence of the capacitance computed as a function of the applied voltage and (a) $N_{d,\mathrm{AL}}$, (b) $N_{d,\mathrm{BL}}$, (c) $N_{d,\mathrm{CL}}$, and (d) t_{B}.

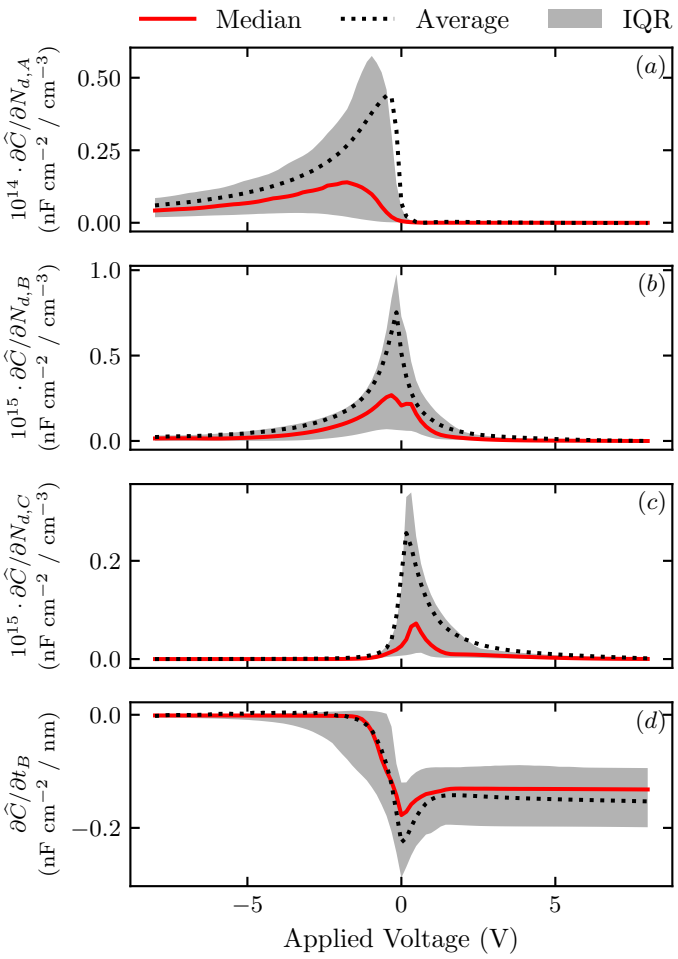

Figure 7·5: The distribution of partial derivatives of the predicted capacitance with respect to (a) $N_{d,\mathrm{AL}}$, (b) $N_{d,\mathrm{BL}}$, (c) $N_{d,\mathrm{CL}}$, and (d) t_{B}.

is repeated for each value of $N_{d,\mathrm{AL}}$ and V_a. The same approach is used to study the other features.

The trends in Fig. 7·4 are the same as what has previously been discussed regarding the role of each feature on the C-V characteristics, but presented in a more concise and visual manner. The partial dependence plots reveals for which voltages and values they offer a significant contribution to determining the capacitance, as well as providing a quantitative description of what we can expect for the average capacitance. For example, Fig. 7·4(a), low values of $N_{d,\mathrm{AL}}$ lead to a rapid decrease in

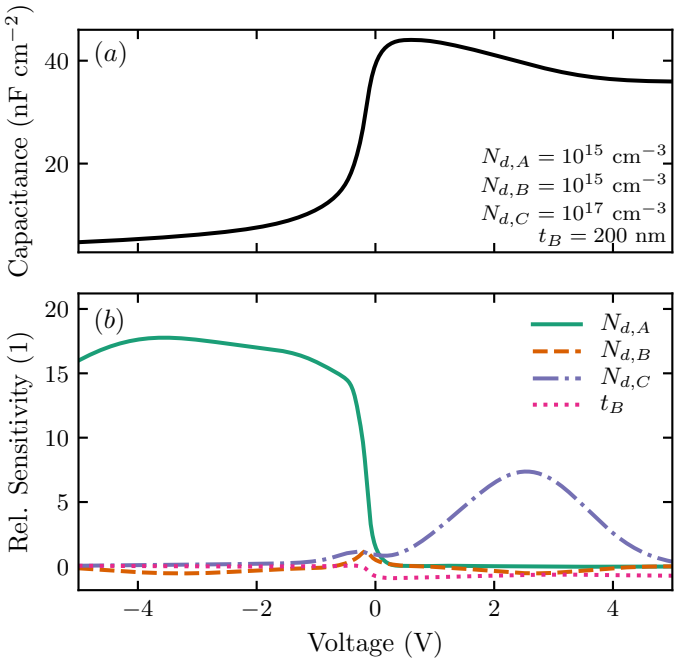

Figure 7·6: (a) Predicted C-V characteristic for the given features. (b) Relative sensitivity of the predicted capacitance to each feature.

capacitance for most cases to a constant value with increasing reverse bias, consistent with a rapidly depleting absorber layer. Under forward bias $N_{d,\mathrm{AL}}$ has negligible effect in determining the value of the capacitance. Fig. 7·4(b) shows the C-V broadening for large values of $N_{d,\mathrm{BL}}$ for smaller biases, and that $N_{d,\mathrm{BL}}$ plays a negligible role in determining the C-V characteristics at extreme values. Figure 7·4(c) highlights the relative invariance of the capacitance with respect to $N_{d,\mathrm{CL}}$, where $N_{d,\mathrm{CL}}$ mostly influences the C-V for moderately high values under forward bias and otherwise has little effect; again, this is due to the thinness of the layer, and rapidly depleting under forward bias for lower doping densities. Finally, Fig. 7·4 highlights that t_{B} is actually quite important in shaping the C-V profile, especially under forward bias. Since the contact layer is thin, and is prone to fully depleting under moderate forward bias, the parallel plate barrier capacitance is important in this regime. For thick barriers, it is also important under low reverse bias in cases where the absorber is either still

accumulated or slightly depleted.

Another common approach to studying the role of the features is to use a gradient-based sensitivity analysis. With automatic differentiation it is trivial to access the partial derivative of the predicted capacitance with respect to any of the features. To gain insight into the global sensitivity of the capacitance to each feature, we compute the partial derivative with respect to each figure for a set of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B} containing the permutation of discrete values within their respective ranges over bias. The distribution of these partial derivatives is shown in Fig. 7·5. The trends are the same as previously discussed: incremental changes in $N_{d,\mathrm{AL}}$ tend to have the largest impact at low reverse bias, $N_{d,\mathrm{BL}}$ at low reverse and forward biases, $N_{d,\mathrm{CL}}$ at low forward bias, and increases in t_{B} will lower the capacitance over much of the voltage range.

Of course, we can also choose a specific design with particular values of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B} and see which the capacitance is most sensitive to. In Fig. 7·6 we directly compare the features by computing a relative sensitivity

$$S_i^* = \frac{\partial \hat{C}}{\partial x} \cdot \frac{x}{C}$$

for $x \in \{N_{d,\mathrm{AL}}, N_{d,\mathrm{BL}}, N_{d,\mathrm{CL}}, t_{\mathrm{B}}\}$. Figure 7·6(b) clearly shows that an incremental change in $N_{d,\mathrm{AL}}$ would have the largest affect in lowering the capacitance, but also provides information on the other features as well. Naturally, this only captures the incremental change in each parameter for this specific set of parameters; for example, increasing $N_{d,\mathrm{BL}}$ by several orders of magnitude would drastically alter the shape of the C-V profile.

7.3 C-V analysis using convolutional neural networks

The previous section dealt with a forward process, where given a set of parameters the capacitance is calculated. In this section, we discuss the inverse, where provided a C-V profile the architectural properties of a device are given. Conventional C-V analysis is often limited in its applicability, requiring strict assumptions about the device structure and only providing information on a few aspects of the device. In practice, a separate witness sample may be required to characterize a critical layer in the device (Rhiger et al., 2016). Numerical techniques can reveal a much deeper perspective into the device, but can be costly to run with a larger barrier of entry. Machine learning offers a way to alleviate some of these constraints. By using data, generated experimentally or numerically, a model can be created that can be used as an enhanced C-V analysis tool that yields more information about the underlying device structure.

Convolutional neural networks (CNNs) are used ubiquitously for signal analysis and have achieved state-of-the-art performance in pattern recognition (LeCun et al., 2015). The widespread use of CNNs for these types of tasks stems from their ability to learn the correspondence between adjacent data points in structured data (Long et al., 2014). Additionally, CNNs are more efficient than fully-connected networks when dealing with larger inputs due to their use of shared weights and biases when processing the input. The sets of weights and biases are referred to as filters, and are used to perform convolutional operations across adjacent data points. The outputs form feature maps that can highlight pertinent information; for example, feature maps have been shown to emphasize certain patterns in images, such as the edges of objects (LeCun et al., 2015, Zeiler and Fergus, 2014, Yosinski et al., 2015).

To illustrate the operation of a CNN, consider the following example. Let I_{ij} represent the intensity of a grayscale pixel located in row i and column j of an image.

Let f be a filter with m rows and n columns. Then, the resulting value in the feature map while considering this position is given by

$$z_{ij}^{f} = \sigma \left[\sum_{q=0}^{m} \sum_{r=0}^{n} \left(w_{qr}^{f} I_{i+q,j+m} \right) + b^{f} \right],$$

where w, b, and σ denote the weights, biases, and activation function. The complete feature map contains the output for every position on the image that the convolution is applied, and contains $n \times m$ weights and a single bias. Unlike the previously discussed fully-connected networks, the number of free parameter has been decoupled from the dimensionality of the preceding layer. The total dimensionality of the filter is determined by the chosen filter size and how it is stepped across the image; these are hyperparameters that must be tuned for each application.

While the number of free parameters is not reliant on the number of connections from the previous layer, the dimensionality can still be large enough that the feature maps are large. Usually, the output of a set of convolutional layers is flattened and given to a smaller fully-connected network for final processing, so it can be useful to further reduce the size of the feature maps. One common approach is to use pooling operations where an assumption is made regarding a cluster of pixels. By taking the minimum, maximum, or average of a window of pixels the size of the feature map can be reduced. Again, the pooling operation, pooling size or number of pixels to include, and where in the image pixels are pooled are all hyperparameters that need to be carefully assessed.

While this brief discussion has been limited to images, CNNs can be useful for any structured data where information is embedded in clusters of points. For this reason, CNNs are researched for use in many areas of signal processing. In the medical industry, for example, CNNs are being used to create models to help healthcare workers in forming diagnoses by analyzing echocardiograms and electroencephalograms (Ki-

ranyaz et al., 2015, Acharya et al., 2017, Acharya et al., 2018). Here, we consider a similar application where the capacitance-voltage characteristics contain information regarding the device architecture.

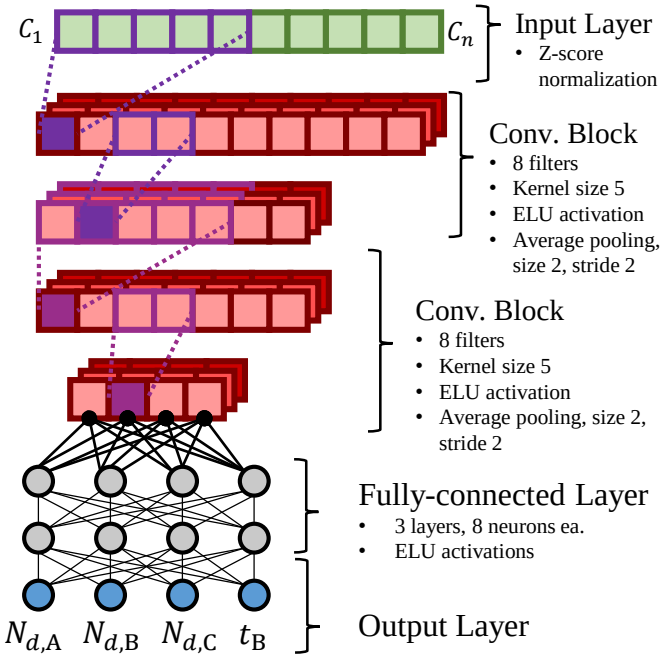

Figure 7·7: The convolutional neural network architecture used in this work to analyze C-V characteristics.

The proposed CNN for predicting $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B} from a C-V profile is shown in Fig. 7·7. The input layer is a one-dimensional array containing the capacitance values.[8] The input is fed to two convolutional blocks, each using eight 5×1 filters, with ELU activation functions. Average pooling was used to reduce the output dimensionality in half by setting the stride and step to two. A global average pooling is applied to the output of the last convolutional block, and the result is given to a fully-connected network with three layers with eight neurons each with ELU activations. The final output is four neurons, representing the predicted $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$,

[8]Since all of the C-V profiles considered in this work were at the same biases, it was not necessary to include the voltage. Voltage could be included as an additional channel if desired. This would, however, increase the number of free parameters, so more data may be required.

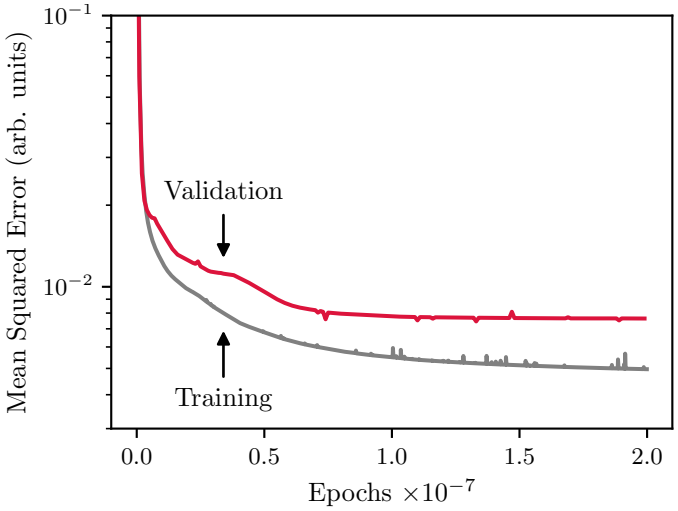

Figure 7·8: Learning curves for a convolutional neural network predicting $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B} from a C-V profile.

$N_{d,\mathrm{CL}}$, and t_{B}. The Adam optimization algorithm was also used for this network with default hyperparameters. Z-score normalize was applied to the inputs and outputs prior to evaluation and training, with the base-10 logarithmic values of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B} being used. A total of 768 quasi-random cases were simulated with 90% used for training and 10% for validation. To improve the model's performance, the 16 values on the corner of the 4-dimensional hypercube were included, and weighted higher during training. The resulting learning curves are shown in Fig. 7·8. Since the complexity of the model was reduced to accommodate the limited number of training samples, the validation error follows the training error without overfitting.

To highlight the effectiveness of the model, predicted-expected curves and the corresponding distribution of their ratios are shown in Fig. 7·9, and a summary of the fitting statistics are presented in Table 7.1. The absorber doping density is accurately predicted over the full doping range with all cases within a factor of two of the expected value. The same is shown to be true for the barrier and contact layer doping densities. The prediction of the barrier thickness is also precise; with the worst case potentially underestimating by about 25 nm, most cases are within 5-15 nm.

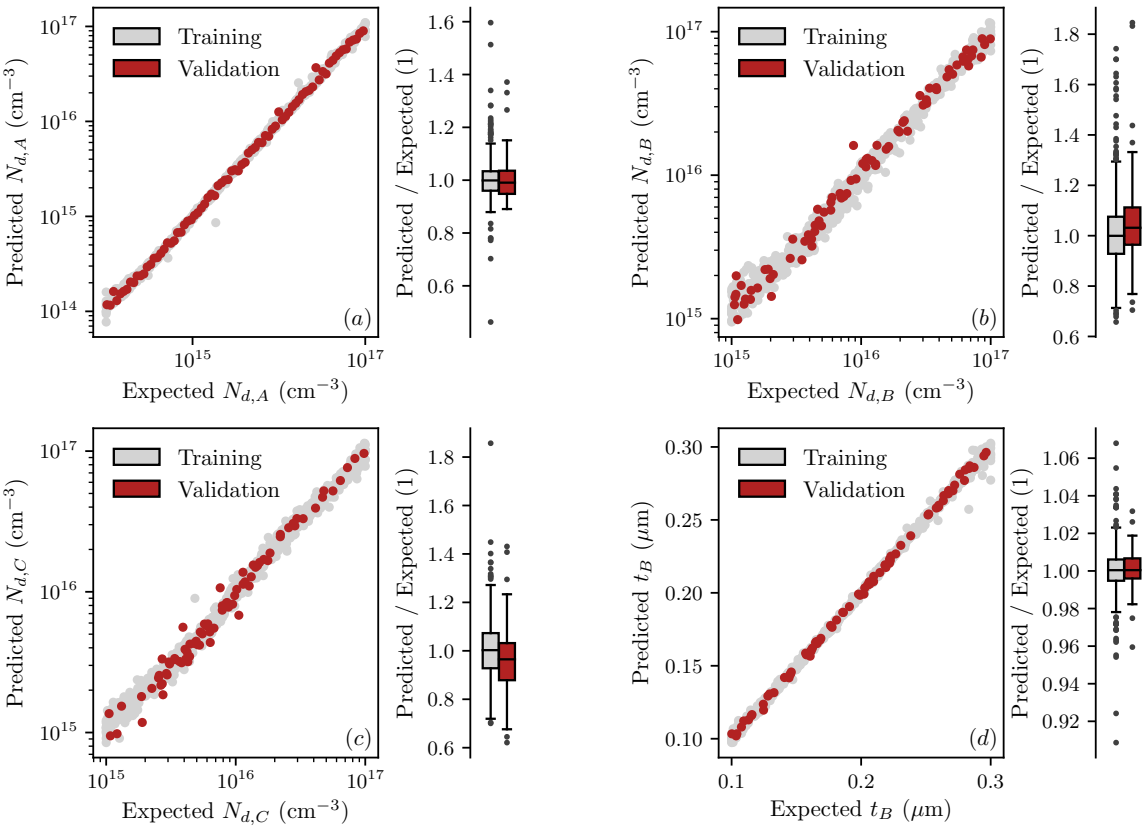

Figure 7·9: Predicted verus expected and their associated distribution of ratios for (a) $N_{d,\mathrm{AL}}$, (b) $N_{d,\mathrm{BL}}$, (c) $N_{d,\mathrm{CL}}$, and (d) t_{B}.

Feature	Dataset	Avg. % Err.	Median % Err.	R^2
$N_{d,\mathrm{AL}}$	Training	-0.273	0.100	0.9931
	Validation	-0.123	0.941	0.9941
$N_{d,\mathrm{BL}}$	Training	-0.940	0.052	0.9815
	Validation	-5.830	-3.145	0.9797
$N_{d,\mathrm{CL}}$	Training	-0.650	-0.275	0.9838
	Validation	3.303	3.510	0.9930
t_{B}	Training	-0.015	-0.041	0.9982
	Validation	-0.055	-0.043	0.9991

Table 7.1: A summary of the achieved performance of a trained convolutional neural network used when predicting $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B} from a C-V profile.

Figure 7·9 also provides insight into which cases are more difficult to predict. Small $N_{d,\mathrm{CL}}$ and large t_{B} tend to be the most inaccurate. The former can be explained by how low $N_{d,\mathrm{CL}}$ leads to the capacitance quickly converging to a constant value, and the latter can be explained by the larger barrier thickness lowering the capacitance over the whole voltage range, diminishing the important aspects of the C-V profile that contain the pertinent information for determining the value of these parameters. Moreover, recalling the discussion of the N-type barrier from the previous chapter, there are cases where the barrier is not fully depleted that cause the general shape of the C-V profile to change dramatically. By including additional samples in these regions of the parameter space, or weighting the available data higher during training would improve the model's accuracy for these cases.

7.4 Surrogate modeling of dark current and quantum efficiency in nBn devices

While capacitance-voltage characteristics are a useful figure of merit for determining the bandwidth of photodetectors and for characterizing semiconducting layers within a device, it is not as commonly used as a figure of merit for comparing infrared sensors as dark current or quantum efficiency. Therefore, it is useful to explore how artificial

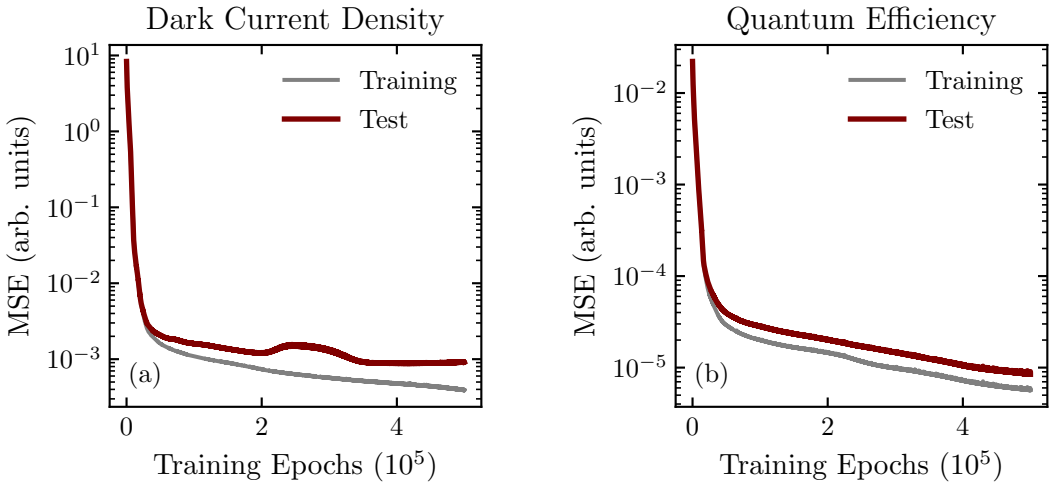

Figure 7·10: Training and test learning curves for (a) dark current and (b) quantum efficiency.

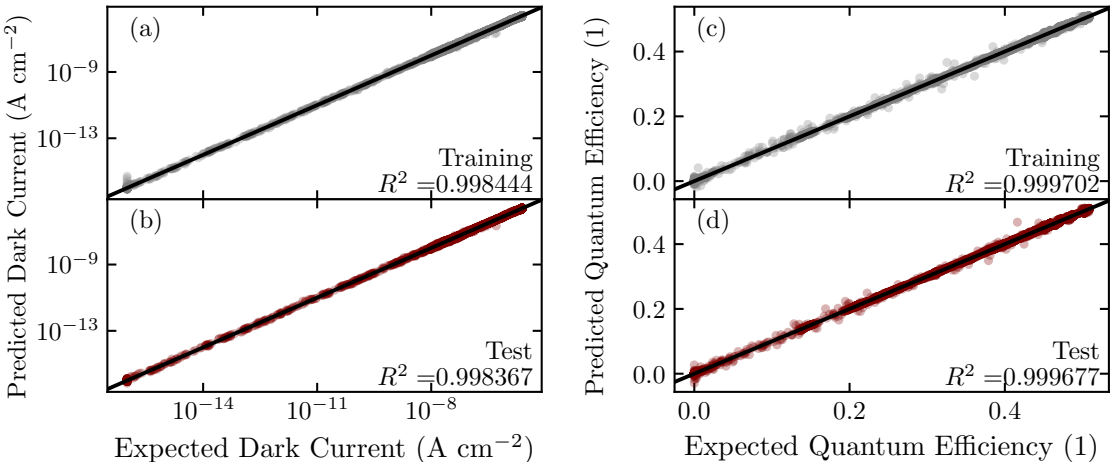

Figure 7·11: Predicted versus expected curves for the (a) and (b) dark current and (c) and (d) quantum efficiency.

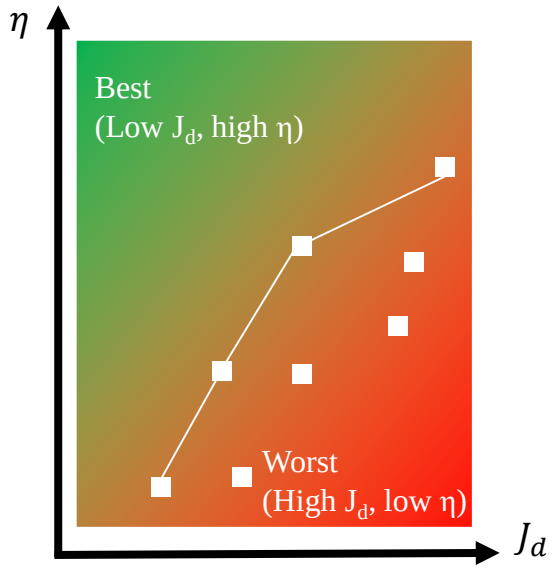

Figure 7·12: An example of a Pareto frontier in terms of quantum efficiency and dark current.

neural networks can be used to explore these metrics. Using the same neural network architecture, methodology, activation functions, and predictors as was discussed for predicting the capacitance as a function of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, t_{B}, and V_a, we trained two separate networks to predict the dark current and QE. The temperature was fixed at 150K, and data was generated at 101 fixed voltages between -8 and 8 V for quasi-random sampled values of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B}. The training results are shown in Figs. 7·10 and 7·11. In both cases, the training and test errors are low without overfitting the data.

A widespread problem in multi-objective optimization is best summarized by the concept of Pareto optimality. Figure 7·12 illustrates the concept in terms of dark current and quantum efficiency. The ideal photodetector would have minimal dark current while having perfect quantum efficiency. Of course this is an unrealistic expectation; each design will have dark current and likely a non-perfect QE. Variations in the design parameters can affect the two in a specific way. Changing the absorber doping density, for example, may lead to a higher QE but also lead to higher dark

current. The sets of design parameters that lead to the highest values of the QE and lowest values of dark current form the Pareto frontier, illustrated by the connected points in Fig. 7·12. This type of analysis provides useful insight into the global parameter design space and the realistic values of QE and dark current that are achievable. By using neural networks to create surrogate models for the dark current and QE, we can fill in the missing data for competing choices of device parameters without additional computational burden.

Shown in Fig. 7·13 is a significant correlation between QE and dark current for $N_{d,\mathrm{AL}}$ within 10^{14}–10^{17} cm^{-3}, $N_{d,\mathrm{BL}}$ within 10^{15}–$10^{17}\mathrm{cm}^{-3}$, $N_{d,\mathrm{CL}}$ within 10^{15}–10^{17} cm^{-3}, and t_{B} within 100–300 nm at 150K and -320 mV. Designs leading to low dark current often have low QE, while some cases that have high QE lead to excess dark current. The green rectangle indicates a region with near optimal characteristics, a high QE and moderately low dark current. The distributions of each feature found in a design that satisfies this criteria is shown in Fig. 7·13(b)-(e), providing invaluable insight into the best choices of each feature to use in practice, as well as the underlying physics. For example, Fig. 7·13(b) highlights that a moderate doping in the absorber is preferred. This is due to low and high absorber doping levels being filtered out due to higher dark current from excess GR current and lower QE due to enhanced Auger recombination. Figures 7·13(c) and (e) show that a high doped, thick barrier layer does not fit the criteria. This is due to how these features affect the crossover voltage where the absorber-barrier interface switches from accumulation to depletion. If the surface is accumualted, there is a valence band barrier that prevents the collection of photogenerated holes, lowering the QE. A higher doping density and thickness lead to a higher crossover voltage, hence, at this relatively low reverse bias the surface will be accumulated and suffer from a lower QE. Shown in Fig. 7·13(d), the contact doping density has no correlation in determining the dark current or QE

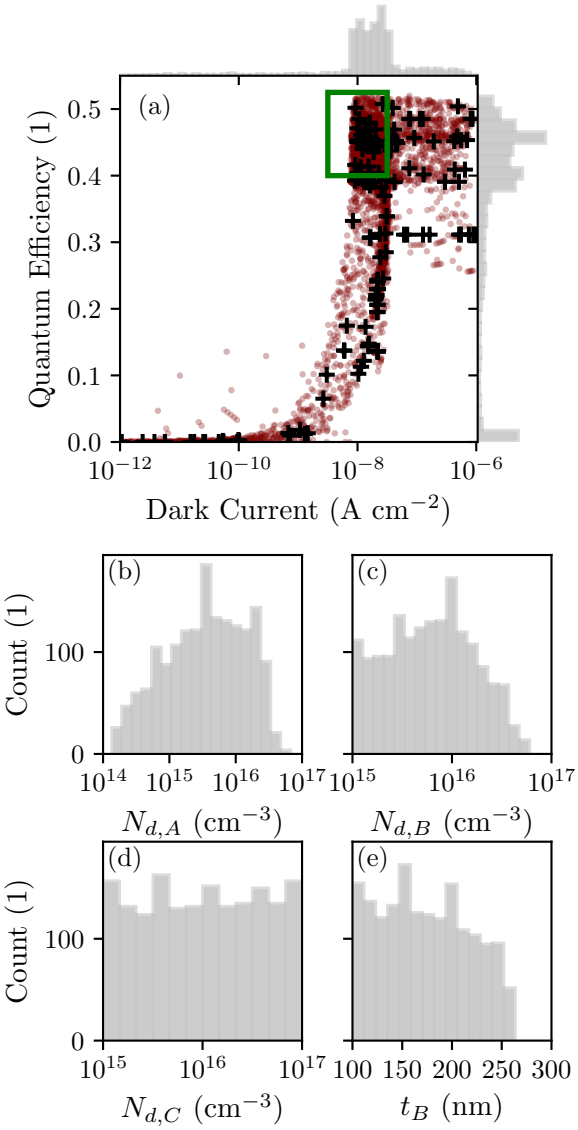

Figure 7·13: (a) Quantum efficiency versus dark current for competing design parameters. The black pluses are simulation points used for training, and the red circles are neural network predictions. (b)-(e) The distributions of features that lead to the best designs highlighted in the green box in (a). $T = 150$K, $V_a = -320$ mV.

under reverse bias.

The trained neural networks can also be used to assess the tolerance of a design to variation in the device parameters. The doping density, or thickness of the layer are known to some accuracy, and will have an experimental tolerance associated with it. For example, consider Fig. 7·14 where we assume that each feature has a Gaussian probability density function. In this case, $N_{d,\mathrm{AL}}$ is centered around 2×10^{15} cm^{-3}, $N_{d,\mathrm{BL}}$ around 5×10^{15} cm^{-3}, $N_{d,\mathrm{CL}}$ around 5×10^{16} cm^{-3}, and t_{B} around 200 nm with standard deviations of 5×10^{14} cm^{-3}, 5×10^{14} cm^{-3}, 10^{16} cm^{-3}, and 25 nm respectively. According to these probability density functions we can sample designs and predict the spread in dark current and quantum efficiency that we can expect. The result is shown in Fig. 7·14(e). For this set of distributions in the features, the spread is quite low; the QE has a slight tendency to lower by a few percent while the dark current increases by around a factor of two.

By increasing the standard deviations of the Gaussian probability density functions, we can study which parameter leads to the highest spread in the figures of merit to optimize the fabrication process or identify which may be leading to lower than expected yield. For example, in Fig. 7·15, the standard deviation of t_{B} has been doubled to 50 nm. This leads to a wide variation in both QE and dark current in Fig. 7·15. Now, the QE varies by $\pm 10\%$ and the dark current by up to a factor of three.

7.5 Conclusion

In this chapter we covered three applications of neural networks to the characteristics of nBn photodetectors with N-type barrier layers. First, we demonstrated that ANNs can accurately reproduce the role of $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, t_{B}, and V_a in determining the C-V characteristics, and can generalize well to cases within the limiting values of

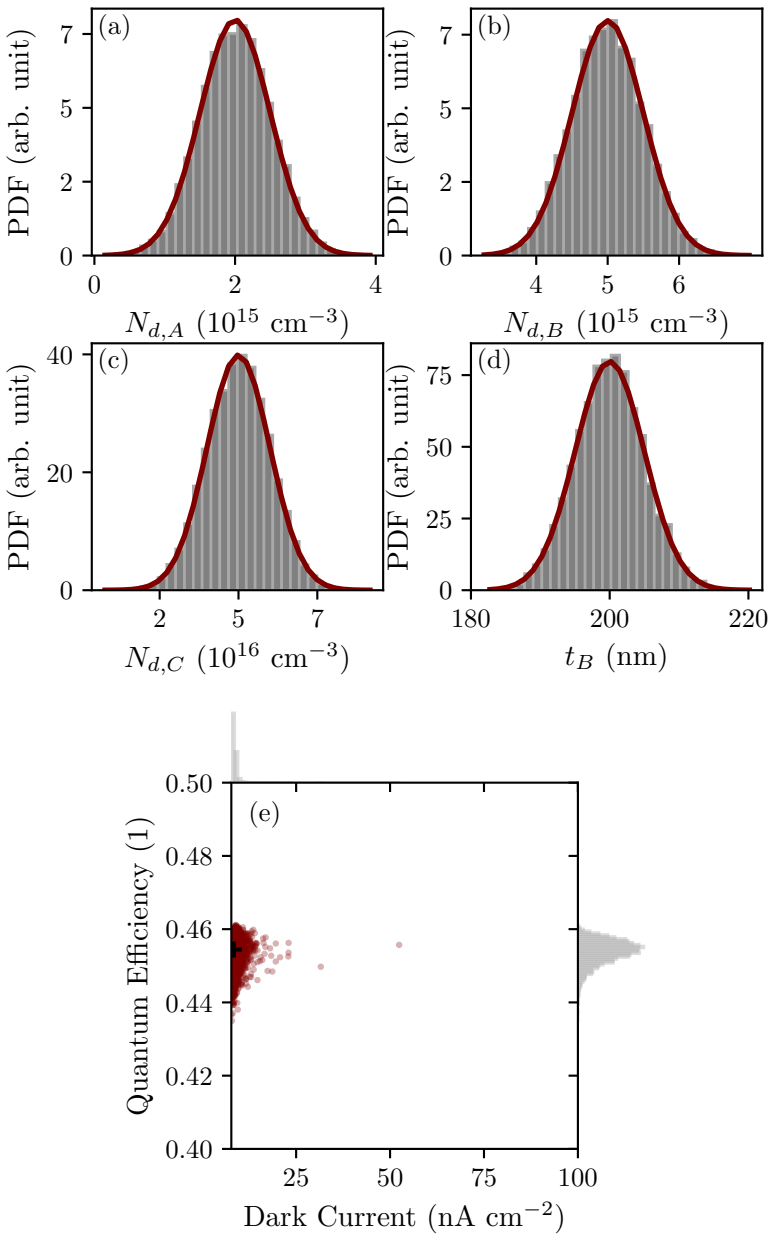

Figure 7·14: Probability distribution function for (a) $N_{d,\mathrm{AL}}$, (b) $N_{d,\mathrm{BL}}$, (c) $N_{d,\mathrm{CL}}$, and (d) t_B. By sampling according to these probability distributions, we assess the spread in quantum efficiency and dark current in (e).

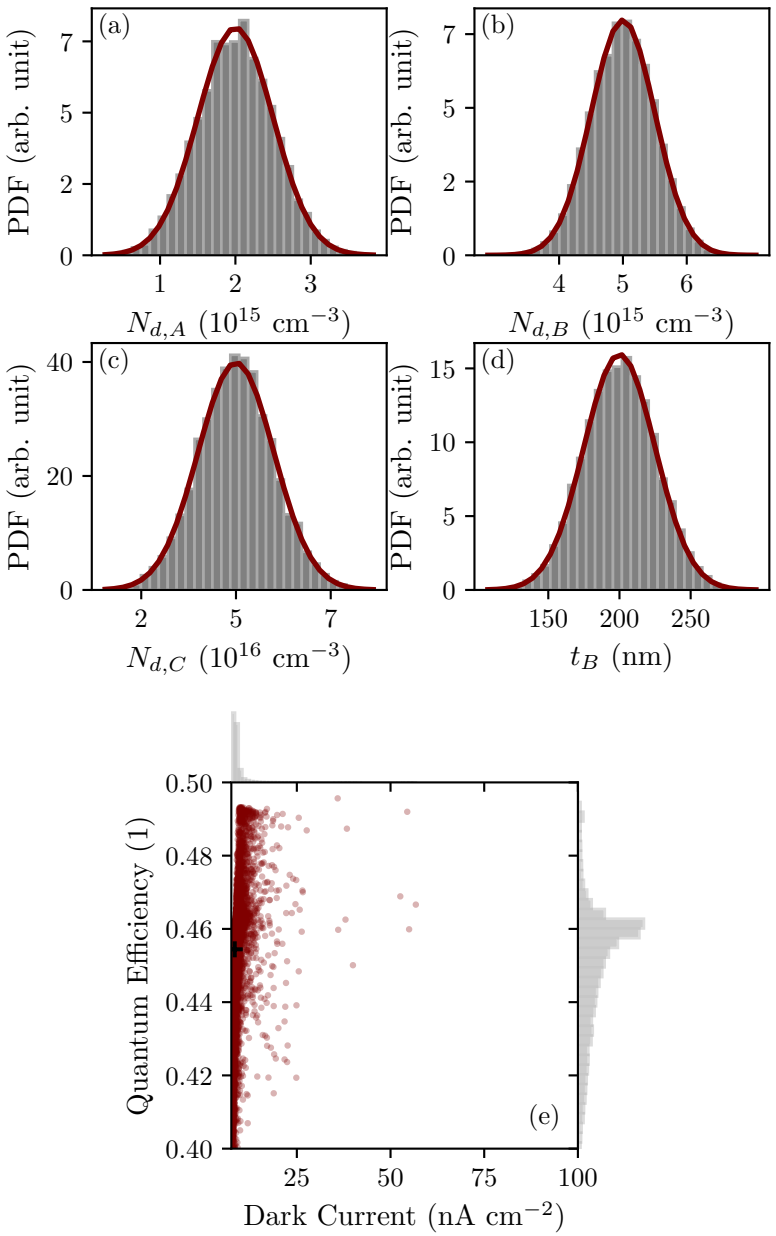

Figure 7·15: Probability distribution function for (a) $N_{d,\mathrm{AL}}$, (b) $N_{d,\mathrm{BL}}$, (c) $N_{d,\mathrm{CL}}$, and (d) t_{B}. By sampling according to these probability distributions, we assess the spread in quantum efficiency and dark current in (e). Note, the distribution of t_{B} has been increased when compared to Fig. 7·14

each feature. ANNs were shown to be a useful in exploring how each feature affects the C-V profile. Second, we established that CNNs can be used to create a model for extracting a comprehensive picture of the underlying device architecture by predicting $N_{d,\mathrm{AL}}$, $N_{d,\mathrm{BL}}$, $N_{d,\mathrm{CL}}$, and t_{B} from a single C-V profile with high accuracy. The CNN can be trained to yield more information than would be possible through conventional C-V analysis. We also applied ANNs to the dark current and quantum efficiency of nBn devices for the same features. Using the ANNs to fill in the feature-space, the tradeoff between QE and dark current was identified, and the best performing devices where shown to have moderate absorber doping densities, low barrier doping densities, and relatively thin barriers; the contact layer doping had no impact under small reverse bias. Lastly, we studied how uncertainties in the features can lead to a spread of quantum efficiency and dark current. To illustrate, a large uncertainty in the barrier thickness was shown to lead to up to 10% variation in quantum efficiency and up to three times increase in dark current.

Chapter 8

Summary of this research

In the first half of this dissertation we discussed the implications of moving to small pitch $In_{0.53}Ga_{0.47}As$ focal plane arrays for SWIR sensing in terms of dark current, quantum efficiency, specific detectivity, and modulation transfer function. Through solutions of the one-dimensional continuity equation for minority carriers, we explored how boundary conditions in planar arrays of photodiodes affect these figures of merit. Specifically it was noted that the Neumann zero flux boundary condition induced by a large valence band barrier to holes in an p-n-N junction improves the quantum efficiency by up to 70%. Additionally, once the pixel pitch is reduced to near the minority carrier diffusion length the neighboring p-n junctions begin to interact by suppressing the minority carrier concentration in the quasi-neutral region between them. This results in a maximum minority carrier concentration halfway between adjacent pixels, and a suppressed lateral diffusion current. By solving the continuity equation in a cylindrical coordinate system, a pseudo-three-dimensional analytical model was shown to be in agreement with a representative three-dimensional drift-diffusion simulation. From this, it was shown that a significant portion of the dark current in planar photodiode arrays is due to the lateral diffusion of minority carriers from the pixel boundaries, even when the quasi-neutral width is reduced below the minority carrier diffusion length.

Using a numerical model based on the drift-diffusion and finite-difference time-domain methods, we studied small-pitch $In_{0.53}Ga_{0.47}As$ focal plane arrays and calcu-

lated the dark current, quantum efficiency, and modulation transfer function. The model was shown to be in agreement with experimental data gathered from literature. The results indicate that while the total dark current decreases with decreasing pitch, due to the suppressed lateral diffusion current, the dark current density rises since for fixed junction radius and absorber thickness, the vertical diffusion current and GR current are constant. While the QE rises slightly with decreasing pitch as less minority carriers recombine before being collected by the junction, the specific detectivity decreases marginally due to the higher dark current density, by less than a factor of two. It was shown that while the MTF improves by moving to smaller-pitch, the sensor's MTF is significantly lower than would be expected by the detector footprint. Isolating the individual components of the MTF showed that the MTF is limited by inter-pixel crosstalk from carriers diffusing into adjacent pixels.

To this end, we offered a proposed change to the pixel sub-architecture with the intent of lowering dark current and mitigating carrier crosstalk by adding additional diffusion control junctions to the pixel corners. For a 10 µm pitch $In_{0.53}Ga_{0.47}As$ photodiode, the dark current was shown to decrease by about a factor of two due to the suppressed minority carrier profile throughout the diode. The photocurrent was also affected by the inclusion of DCJs into the structure, leading to the QE decreasing by 20%; accordingly, the specific detectivity was also lowered, but by less than a factor of two when compared to the baseline structure without DCJs. However, the spot-scan profile exhibited a significant reduction in crosstalk when the Gaussian beam is incident on adjacent pixels, leading to an improved MTF. At the Nyquist frequency the MTF was shown to improve from 0.37 to 0.5. With the goal of further reducing the dark current we also investigated an annular DCJ to exploit radial symmetry to further suppress the minority carrier profile. This was shown to be effective at reducing the dark current, achieving close to an order of

magnitude reduction. However, the photocurrent is again negatively affected, leading to a severely degraded QE by more than 60%. Unfortunately, the reduction in dark current is not enough to improve D^*, which is lower by a factor of three. Finally, it was shown that by coupling the DCJ approach with this annular structure with a form of optical generation localization—in this case we used monolithically integrated microlenses—the DCJ approach can offer an estimated improvement from 2.75×10^{13} to 3.75×10^{13} Jones, due to mitigating the QE loss.

Continuing the development of approaches of simulating infrared detector figures of merit, we developed a methodology for calculating the capacitance-voltage characteristics of nBn devices. First, a semi-analytical model based on MOS theory was shown to agree well with a drift-diffusion model, and was used to fit data available in the literature, providing more details about the underlying device architecture than what is typically given by conventional C-V profiling. Using a one-dimensional drift-diffusion model we performed a comprehensive parametric study to understand the role of the absorber, barrier, and contact doping density, and barrier thickness in determining the C-V characteristics of nBn devices with N- and P-type barriers. For low N-type barriers, the C-V characteristics were shown to be largely determined by the junction capacitances from the depletions regions within the absorber and contact layers under reverse and forward biases respectively. As the N-type doping density is increased, the crossover voltages where the absorber-barrier and barrier-contact interfaces switch from accumulation to depletion increases, resulting in a broadening of the C-V profile at low-moderate biases. The barrier thickness, while the barrier is fully depleted, imposes a limit on the C-V near equilibrium due to the parallel plate capacitance dominating. For large doping densities the assumption of a fully depleted barrier breaks down, where the barrier is comprised of two disparate depletion regions separated by a thin quasi-neutral region. This results in the capacitance exceeding

the parallel plate capacitance limit, as the two depletion capacitances in series led to a larger total capacitance. The case of a P-type barrier was shown to be similar to the N-type for low doping levels, with the exception that the total capacitance is lower near equilibrium as both the absorber-barrier and contact-barrier interfaces are depleted. Large P-type doping densities led to unique C-V curves, which were explained by considering the interplay between the absorber depletion region, the barrier depletion regions while not fully depleted, and the contact layer depletion region under bias.

Last, we applied artificial neural networks to nBn device characteristics. Beginning with the aforementioned C-V characteristics, we demonstrated that ANNs could be used to reduce the number of calculations required to map the parameter space. Using partial density plots and gradient-based sensitivity analyses, the important aspects of the previously discussed semiconductor device physics are elucidated by the model. We also investigated the inverse case, where a C-V profile is given as an input to a convolutional neural network. The CNN then predicts the device architectural parameters. In this work, we included the absorber, barrier, and contact doping densities, and the barrier thickness. The CNN was shown to give accurate results over a wide range of values for each feature, demonstrating the efficacy of using a CNN as an enhanced C-V profiling method. We also applied ANNs to the dark current and quantum efficiency of nBn devices to explore the tradeoffs and map the parameter space for the same set of features used during the C-V study. It was shown that ANNs are useful as a surrogate model for these figures of merit to map the multi-dimensional parameter space with less computational burden. The models revealed that designs with the highest quantum efficiency and low dark current tend to have a moderate doping density in the absorber and a relatively thin low doped barrier layer; the contact layer has no correlation to either figure of merit under reverse bias. We

also showed how ANNs could be used to quantify the spread of quantum efficiency in dark current due to uncertainties in each feature; for example, an uncertainty in barrier thickness of about ± 25 nm led to the quantum efficiency varying by up to 10%, and the dark current increasing by a factor of three.

Appendix A

Material Parameters

The material parameters used when performing the calculations in this thesis have been adapted from previous works (Wichman et al., 2014, Schuster et al., 2014). The models and references are summarized here for convenience.

A.1 $In_{0.53}Ga_{0.47}As$

The bandgap of $In_{0.53}Ga_{0.47}As$ is modeled by (Zielinski et al., 1986, Vurgaftman et al., 2001)

$$E_g(T) = 0.731 - 3.34 \times 10^{-4}(T - 300), \tag{A.1}$$

where T is the temperature in Kelvin. The electron affinity is given by (IOFFE, 2007, Vurgaftman et al., 2001)

$$\chi(T) = \chi_0 + \frac{\alpha T^2}{2(T + \beta)}, \tag{A.2}$$

where $\chi_0 = 4.54716$, $\alpha = 4.3158 \times 10^{-4}$ eV K^{-2}, and $\beta = 227.31$ eV K^{-1}. The conduction and valence band density of states are (Adachi, 1992, Vurgaftman et al., 2001)

$$N_c(T) = 2.08328 \times 10^{18} \left(\frac{T}{300}\right)^{3/2}, \tag{A.3}$$

$$N_v(T) = 7.8137 \times 10^{18} \left(\frac{T}{300}\right)^{3/2}, \tag{A.4}$$

respectively. The radiative and Auger coefficients are (Ahrenkiel et al., 1998)

$$B(T) = 9.5 \times 10^{11} \left(\frac{T}{300}\right)^{-1.68} \tag{A.5}$$

$$C_{n,p} = 8.1 \times 10^{-29}. \tag{A.6}$$

The SRH lifetime, τ_{SRH}, was fixed at 107 μs (DeWames et al., 2015). The electron and hole mobilities are (Madelung, 2012)

$$\mu_e = 1.24 \times 10^4 \left(\frac{T}{300}\right)^{-1.405}, \tag{A.7}$$

$$\mu_h = 234 \left(\frac{T}{300}\right)^{-1.42}, \tag{A.8}$$

respectively. The relative dielectric constant was set to $\epsilon_r = 13.9061$ (IOFFE, 2007). Lastly, the data for the index of refraction, n, and extinction coefficient, k, were provided by (Muñoz et al., 2002, Madelung, 2012).

A.2 InP

The bandgap and electron affinity of InP are (Vurgaftman et al., 2001, Adachi, 1992)

$$E_g(T) = E_{g,0} + \frac{\alpha T^2}{T + \beta}, \tag{A.9}$$

$$\chi(T) = \chi_0 + \frac{\alpha T^2}{T + \beta}, \tag{A.10}$$

respectively where $\alpha = 4.1 \times 10^{-4}$ eV K^{-2}, $\beta = 136$ eV K^{-1}, $E_{g,0} = 1.4205$ eV, and $\chi_0 = 4.4$ eV. The conduction and valence band density of states are (IOFFE, 2007, Adachi, 1992)

$$N_c(T) = 5.66 \times 10^{17} \left(\frac{T}{300}\right)^{3/2}, \tag{A.11}$$

$$N_v(T) = 2.03 \times 10^{19} \left(\frac{T}{300}\right)^{3/2}. \tag{A.12}$$

The electron and hole mobilities are (IOFFE, 2007, Adachi, 1992)

$$\mu_e = 4500 \left(\frac{T}{300} \right)^{1.5}, \tag{A.13}$$

$$\mu_h = 150 \left(\frac{T}{300} \right)^{1.5}. \tag{A.14}$$

The relative dielectric constant is $\epsilon_r = 12.4$ (IOFFE, 2007). The data for the index of refraction, n, and extinction coefficient, k, are found in (Adachi, 1989). The radiative and Auger recombination are assumed to be negligible since InP has a relatively large bandgap compared to $In_{0.53}Ga_{0.47}As$; the constants are arbitrarily set to 0 $cm^3\,s^{-1}$ and 10^{-30} $cm^6\,s^{-1}$.

A.3 $InAs_{1-x}Sb_x$

As mentioned previously, the materials parameters for $InAs_{1-x}Sb_x$ used in this dissertation were adopted from (Schuster et al., 2014).

The bandgap for arbitrary composition and temperature for $InAs_{1-x}Sb_x$ is often described by (Vurgaftman et al., 2001, Adachi, 1987)

$$E_g(x,T) = 0.411 - \frac{3.4 \times 10^{-4} T^2}{210 + T} - 0.876x + 0.7x^2 + 3.4 \times 10^{-3} xT(1-x). \tag{A.15}$$

The electron affinity is an interpolation between InAs and InSb with a bowing parameter:

$$\chi_{\mathrm{InAs}} = 5.06 + \frac{0.000276 T^2}{2(93+T)} \tag{A.16}$$

$$\chi_{\mathrm{InSb}} = 4.72 + \frac{0.00032 T^2}{2(170+T)} \tag{A.17}$$

$$\chi(x,T) = (1-x)\chi_{\mathrm{InAs}} + x\chi_{\mathrm{InSb}} - 0.6x(1-x) \tag{A.18}$$

The electron and hole effective masses along with the conduction and valence band

density of states are modeled by (Vurgaftman et al., 2001, IOFFE, 2007)

$$m_e^*(x) = m_0(0.023 - 0.039x + 0.03x^2), \tag{A.19}$$

$$m_h^*(x) = m_0\left[(0.026 - 0.011x)^{3/2} + (0.41 + 0.02x)^{3/2}\right]^{2/3}, \tag{A.20}$$

$$N_c(x,T) = 2.54 \times 10^{19}\left(\frac{m_e^*(x)}{m_0}\cdot\frac{T}{300}\right)^{3/2}, \tag{A.21}$$

$$N_v(x,T) = 2.54 \times 10^{19}\left(\frac{m_h^*(x)}{m_0}\cdot\frac{T}{300}\right)^{3/2}, \tag{A.22}$$

where m_0 is the electron rest mass. The electron and hole mobilities are linearly interpolated from the binary constituents, InSb and InAs: (Adachi, 1992)

$$\mu_e = 3\times 10^4(1-x)\left(\frac{T}{300}\right)^{-0.7212} + 8\times 10^4 x\left(\frac{T}{300}\right)^{-1.8572}, \tag{A.23}$$

$$\mu_h = 500(1-x)\left(\frac{T}{300}\right)^{-0.5097} + 800x\left(\frac{T}{300}\right)^{-1.8572}. \tag{A.24}$$

The dielectric constant is also linearly interpolated, $\epsilon_r = 16.8x + 15.15(1-x)$ (IOFFE, 2007). The index of refraction is modeled using an interpolation over composition and wavelength to the data in (Paskov, 1997). The absorption coefficient is calculated using the model from (D'souza et al., 2012):

$$\alpha(x,T,E_{\text{ph}}) = \begin{cases} 948.23\, e^{170(E_{\text{ph}}-E_0)} & E_{\text{ph}} < E_g(x,T) \\ \dfrac{K(E_{\text{ph}}-E_g-c)}{E_{\text{ph}}}\left[(E_{\text{ph}}-E_g-c)^2 - c^2\right]^{1/2} & E_{\text{ph}} \geq E_g(x,T) \end{cases} \tag{A.25}$$

where $K = 20000E_g + 10000$, $c = -0.5E_g - 0.1$, $E_0 = E_g + 0.001$, and $E_{\text{ph}} = 1.24/\lambda$. The radiative coefficient is computed using (Bellotti and D'Orsogna, 2006, Schuster et al., 2014)

$$C_r = 5.8\times 10^{-13}\epsilon_\infty^{\frac{1}{2}}\left(\frac{m_0}{m_e^*+m_h^*}\right)^{\frac{3}{2}}\left(1+\frac{m_0}{m_e^*}+\frac{m_0}{m_v^*}\right)\times$$

$$\left(\frac{300}{T}\right)\left(E_g^2 - 3k_BTE_g + 3.75k_B^2T^2\right). \quad \text{(A.26)}$$

The Auger coefficients are (Bellotti and D'Orsogna, 2006, Schuster et al., 2014)

$$C_n = \frac{\frac{m_e^*}{m_0}|F_1F_2|^2}{2n_i^2 3.8\times10^{-18}\epsilon_\infty^2\left(1+\frac{m_e^*}{m_0}^{1/2}\right)\left(1+2\frac{m_e^*}{m_h^*}\right)}\left(\frac{E_g}{k_BT}^{-\frac{3}{2}}\right)\times \exp\left(-\frac{1+2\frac{m_e^*}{m_h^*}}{1+\frac{m_e^*}{m_h^*}}\frac{E_g}{k_BT}\right), \quad \text{(A.27)}$$

and

$$C_p = C_n\left[\frac{1-\frac{3E_g}{k_BT}}{6\left(1-\frac{5E_g}{4k_BT}\right)}\right]. \quad \text{(A.28)}$$

The overlap integral $|F_1F_2|$ was set to 0.1 (Wen et al., 2015).

References

Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adam, M. (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences*, 415:190–198.

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in biology and medicine*, 100:270–278.

Adachi, S. (1987). Band gaps and refractive indices of AlGaAsSb, GaInAsSb, and InPAsSb: Key properties for a variety of the 2–4-μm optoelectronic device applications. *Journal of applied physics*, 61(10):4869–4876.

Adachi, S. (1989). Optical dispersion relations for GaP, GaAs, GaSb, InP, InAs, InSb, $Al_xGa_{1-x}As$, and $In_{1-x}Ga_xAs_yP_{1-y}$. *Journal of Applied Physics*, 66(12):6030–6040.

Adachi, S. (1992). *Physical properties of III-V semiconductor compounds*. John Wiley & Sons.

Adachi, S. (2009). *Properties of semiconductor alloys: group-IV, III-V and II-VI semiconductors*, volume 28. John Wiley & Sons.

Ahrenkiel, R., Ellingson, R., Johnston, S., and Wanlass, M. (1998). Recombination lifetime of In0.53Ga0.47As as a function of doping density. *Applied Physics Letters*, 72(26):3470–3472.

Bellotti, E. and D'Orsogna, D. (2006). Numerical analysis of HgCdTe simultaneous two-color photovoltaic infrared detectors. *IEEE Journal of Quantum Electronics*, 42(4):418–426.

Boreman, G. D. (2001). *Modulation transfer function in optical and electro-optical systems*, volume 10. SPIE press Bellingham, WA.

Capper, P., Kasap, S., and Willoughby, A. (2011). *Mercury Cadmium Telluride: Growth, Properties and Applications*, volume 38. John Wiley & Sons.

Casey, H. C. (1999). *Devices for Integrated Circuits: Silicon and III-V Compound Semiconductors*. Wiley.

Choo, S. (1968). Carrier generation-recombination in the space-charge region of an asymmetrical pn junction. *Solid-State Electronics*, 11(11):1069–1077.

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*.

DeWames, R., Littleton, R., Witte, K., Wichman, A., Bellotti, E., and Pellegrino, J. (2015). Electro-optical characteristics of p+n ingaas hetero-junction photodiodes in large format dense focal plane arrays. *Journal of Electronic Materials*, 44(8):2813–2822.

Dhar, N. K., Dat, R., and Sood, A. K. (2013). Advances in infrared detector array technology. *Optoelectronics-Advanced Materials and Devices*, 1.

Dolas, M. H., Atesal, O., Caliskan, M. D., Bek, A., and Ozbay, E. (2019). Low dark current diffusion limited planar type InGaAs photodetectors. In *Infrared Sensors, Devices, and Applications IX*, volume 11129 of *Proceedings of SPIE*. International Society for Optics and Photonics.

D'souza, A., Robinson, E., Ionescu, A., Okerlund, D., De Lyon, T., Sharifi, H., Roebuck, M., Yap, D., Rajavel, R., Dhar, N., et al. (2012). Electrooptical characterization of MWIR InAsSb detectors. *Journal of electronic materials*, 41(10):2671–2678.

Ettenberg, M. H., Lange, M. J., Sugg, A. R., Cohen, M. J., and Olsen, G. H. (1999). Zinc diffusion in InAsP/InGaAs heterostructures. *Journal of Electronic Materials*, 28(12):1433–1439.

Fraenkel, R., Aronov, D., Benny, Y., Berkowicz, E., Bykov, L., Calahorra, Z., Fishman, T., Giladi, A., Ilan, E., Klipstein, P., et al. (2012). SCD's cooled and uncooled photo detectors for NIR SWIR. In *Infrared Technology and Applications XXXVIII*, volume 8353 of *Proceedings of SPIE*. International Society for Optics and Photonics.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Glasmann, A., Hubbard, T., and Bellotti, E. (2017). Numerical modeling of a dark current suppression mechanism in ir detector arrays. In *Infrared Technology and Applications XLIII*, volume 10177 of *Proceedings of SPIE*. International Society for Optics and Photonics.

Glasmann, A., Prigozhin, I., and Bellotti, E. (2019). Understanding the C-V characteristics of InAsSb-based nBn infrared detectors with N-and P-Type barrier layers through numerical modeling. *IEEE Journal of the Electron Devices Society*, 7:534–543.

Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Greenwell, B. M. (2017). pdp: an r package for constructing partial dependence plots. *The R Journal*, 9(1):421–436.

Grimbergen, C. (1976). The influence of geometry on the interpretation of the current in epitaxial diodes. *Solid-State Electronics*, 19(12):1033–1037.

Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1):84–90.

Halton, J. H. (1964). Algorithm 247: Radical-inverse quasi-random point sequence. *Communications of the ACM*, 7(12):701–702.

Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). *The elements of statistical learning: data mining, inference and prediction*, chapter 10, pages 369–370. Springer, 2 edition.

Herschel, W. (1800). Xiv. experiments on the refrangibility of the invisible rays of the sun. *Philosophical Transactions of the Royal Society of London*, (90):284–292.

Holst, G. C. (2007). Imaging system performance based upon Fλ/d. *Optical Engineering*, 46(10):103204.

Horn, S., Norton, P., Cincotta, T., Stoltz Jr, A. J., Benson, J., Perconti, P., and Campbell, J. (2003). Challenges for third-generation cooled imagers. In *Infrared Technology and Applications XXIX*, volume 5074 of *Proceedings of SPIE*, pages 44–51. International Society for Optics and Photonics.

Huang, A., Zhong, Z., Wu, W., and Guo, Y. (2016). An artificial neural network-based electrothermal model for gan hemts with dynamic trapping effects consideration. *IEEE Transactions on Microwave Theory and Techniques*, 64(8):2519–2528.

IOFFE (2007). NSM Archive - Physical Properties of Semicondcutors. Online. `http://www.ioffe.ru/SVA/NSM/Semicond/GaInAs/index.html`.

Irani, K. B., Cheng, J., Fayyad, U. M., and Qian, Z. (1993). Applying machine learning to semiconductor manufacturing. *IEEE Expert*, 8(1):41–47.

Jordan, M. and Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260.

Kasap, S. (2006). *Springer handbook of electronic and photonic materials*. Springer Science & Business Media.

Kinch, M. A. (2000). Fundamental physics of infrared detector materials. *Journal of Electronic Materials*, 29(6):809–817.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kiranyaz, S., Ince, T., and Gabbouj, M. (2015). Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3):664–675.

Klipstein, P. (2008). "XBn" barrier photodetectors for high sensitivity and high operating temperature infrared sensors. In *Infrared Technology and Applications XXXIV*, volume 6940 of *Proceedings of SPIE*. International Society for Optics and Photonics.

Klipstein, P., Klin, O., Grossman, S., Snapi, N., Yaakobovitz, B., Brumer, M., Lukomsky, I., Aronov, D., Yassen, M., Yofis, B., et al. (2010). MWIR InAsSb XBn detectors for high operating temperatures. In *Infrared Technology and Applications XXXVI*, volume 7660 of *Proceedings of SPIE*. International Society for Optics and Photonics.

Kocis, L. and Whiten, W. J. (1997). Computational investigations of low-discrepancy sequences. *ACM Transactions on Mathematical Software (TOMS)*, 23(2):266–294.

Laux, S. E. (1985). Techniques for small-signal analysis of semiconductor devices. *IEEE Transactions on Electron Devices*, 32(10):2028–2037.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436.

Lin, Y., Donetsky, D., Wang, D., Westerfeld, D., Kipshidze, G., Shterengas, L., Sarney, W. L., Svensson, S. P., and Belenky, G. (2015). Development of bulk InAsSb alloys and barrier heterostructures for long-wave infrared detectors. *Journal of Electronic Materials*, 44(10):3360–3366.

Litovski, V. B., Radjenovic, J. I., Mrcarica, Z. M., and Milenkovic, S. L. (1992). Mos transistor modelling using neural network. *Electronics Letters*, 28(18):1766–1768.

Long, J. L., Zhang, N., and Darrell, T. (2014). Do convnets learn correspondence? In *Advances in Neural Information Processing Systems*, pages 1601–1609.

MacDougal, M., Hood, A., Geske, J., Wang, J., Patel, F., Follman, D., Manzo, J., and Getty, J. (2011). InGaAs focal plane arrays for low-light-level SWIR imaging. In *Infrared Technology and Applications XXXVII*, volume 8012 of *Proceedings of SPIE*. International Society for Optics and Photonics.

MacDougal, M. H., Hood, A., Geske, J., Tribble, M., and Hewitt, A. (2020). Small pixel VGA SWIR cameras for laser sensing (Conference Presentation). In Andresen, B. F., Fulop, G. F., Miller, J. L., and Zheng, L., editors, *Infrared Technology and Applications XLVI*, volume 11407 of *Proceedings of SPIE*. International Society for Optics and Photonics, SPIE.

Madelung, O. (2012). *Semiconductors: Group IV elements and III-V compounds*. Springer Science & Business Media.

Maimon, S. and Wicks, G. (2006). nBn detector, an infrared detector with reduced dark current and higher operating temperature. *Applied Physics Letters*, 89(15):151109.

Melati, D., Grinberg, Y., Dezfouli, M. K., Janz, S., Cheben, P., Schmid, J. H., Sánchez-Postigo, A., and Xu, D.-X. (2019). Mapping the global design space of nanophotonic components using machine learning pattern recognition. *Nature communications*, 10(1):1–9.

Moyne, J. and Iskandar, J. (2017). Big data analytics for smart manufacturing: Case studies in semiconductor manufacturing. *Processes*, 5(3).

Muller, R. and Kamins, T. (2003). *Device Electronics for Integrated Circuits*. John Wiley and Sons.

Muñoz, M., Holden, T. M., Pollak, F. H., Kahn, M., Ritter, D., Kronik, L., and Cohen, G. M. (2002). Optical constants of $In_{0.53}Ga_{0.47}As$/InP: experiment and modeling. *Journal of applied physics*, 92(10):5878–5885.

Neidinger, R. D. (2010). Introduction to automatic differentiation and matlab object-oriented programming. *SIAM review*, 52(3):545–563.

Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. Determination Press.

Onat, B. M., Huang, W., Masaun, N., Lange, M., Ettenberg, M. H., and Dries, C. (2007). Ultra-low dark current ingaas technology for focal plane arrays for low-light level visible-shortwave infrared imaging. In *Infrared Technology and Applications XXXIII*, volume 6542 of *Proceedings of SPIE*. International Society for Optics and Photonics.

Paskov, P. (1997). Refractive indices of InSb, InAs, GaSb, $InAs_xSb_{1-x}$, and $In_{1-x}Ga_xSb$: effects of free carriers. *Journal of applied physics*, 81(4):1890–1898.

Pinkie, B., Schuster, J., and Bellotti, E. (2013). Physics-based simulation of the modulation transfer function in HgCdTe infrared detector arrays. *Optics letters*, 38(14):2546–2549.

Reine, M., Pinkie, B., Schuster, J., and Bellotti, E. (2014). Numerical simulation and analytical modeling of InAs nBn infrared detectors with n-type barrier layers. *Journal of electronic materials*, 43(8):2915–2934.

Reine, M. B. (2001). HgCdTe photodiodes for IR detection: A review. In *Photodetectors: Materials and Devices VI*, volume 4288 of *Proceedings of SPIE*, pages 266–277. International Society for Optics and Photonics.

Rhiger, D. R., Smith, E. P., Kolasa, B. P., Kim, J. K., Klem, J. F., and Hawkins, S. D. (2016). Analysis of III–V superlattice nBn device characteristics. *Journal of Electronic Materials*, 45(9):4646–4653.

Rogalski, A. (2010). *Infrared detectors*. CRC press.

Rogalski, A. (2012). History of infrared detectors. *Opto-Electronics Review*, 20(3):279–308.

Rogalski, A., Antoszewski, J., and Faraone, L. (2009). Third-generation infrared photodetector arrays. *Journal of applied physics*, 105(9):4.

Rogalski, A., Martyniuk, P., and Kopytko, M. (2016). Challenges of small-pixel infrared detectors: a review. *Reports on Progress in Physics*, 79(4):046501.

Root, D. E. (2012). Future device modeling trends. *IEEE Microwave Magazine*, 13(7):45–59.

Schroder, D. K. (2015). *Semiconductor material and device characterization*. John Wiley & Sons.

Schuster, J., D'Souza, A., and Bellotti, E. (2014). Analysis of InAsSb nBn spectrally filtering photon-trapping structures. *Optics express*, 22(16):18987–19004.

Susto, G. A., Pampuri, S., Schirru, A., Nicolao, G. D., McLoone, S. F., and Beghi, A. (2012). Automatic control and machine learning for semiconductor manufacturing: Review and challenges. In *10th European Workshop on Advanced Control and Diagnosis (ACD 2012)*.

Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., and Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, 11(3):812–820.

Synopsys (2017a). *Sentaurus Device User Guide*. Synopsys.

Synopsys (2017b). *Sentaurus Electromagnetic Wave Solver User Guide*. Synopsys.

Sze, S. M. and Ng, K. K. (2006). *Physics of semiconductor devices*. John wiley & sons.

Taflove, A. and Hagness, S. C. (2005). *Computational electrodynamics: the finite-difference time-domain method*. Artech house.

Trezza, J. A., Masaun, N., and Ettenberg, M. (2011). Analytic modeling and explanation of ultra-low noise in dense SWIR detector arrays. In *Infrared Technology and Applications XXXVII*, volume 8012 of *Proceedings of SPIE*. International Society for Optics and Photonics.

Vatsia, M. L. (1972). Atmospheric optical environment. Technical report, Night Vision Laboratory, Fort Belvoir, Virginia. `https://apps.dtic.mil/sti/citations/AD0750610`.

Vincent, J. D., Hodges, S., Vampola, J., Stegall, M., and Pierce, G. (2015). *Fundamentals of Infrared and Visible Detector Operation and Testing*. John Wiley & Sons.

Vurgaftman, I., Meyer, J. á., and Ram-Mohan, L. á. (2001). Band parameters for iii–v compound semiconductors and their alloys. *Journal of applied physics*, 89(11):5815–5875.

Wen, H., Pinkie, B., and Bellotti, E. (2015). Direct and phonon-assisted indirect auger and radiative recombination lifetime in HgCdTe, InAsSb, and InGaAs computed using green's function formalism. *Journal of Applied Physics*, 118(1):015702.

White, A. M. (1987). Infra red detectors. US Patent 4,679,063.

Wichman, A. R., DeWames, R. E., and Bellotti, E. (2014). Three-dimensional numerical simulation of planar P+n heterojunction In0.53Ga0.47As photodiodes in dense arrays part I: Dark current dependence on device geometry. In *SPIE Defense+ Security*, pages 907003–907003. International Society for Optics and Photonics.

Yee, K. (1966). Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Transactions on antennas and propagation*, 14(3):302–307.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., and Lipson, H. (2015). Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579*.

Yuan, H., Meixell, M., Zhang, J., Bey, P., Kimchi, J., and Kilmer, L. C. (2012). Low dark current small pixel large format ingaas 2D photodetector array development at Teledyne Judson Technologies. In *Infrared Technology and Applications XXXVIII*, volume 8353 of *Proceedings of SPIE*. International Society for Optics and Photonics.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer.

Zielinski, E., Schweizer, H., Streubel, K., Eisele, H., and Weimann, G. (1986). Excitonic transitions and exciton damping processes in InGaAs/InP. *Journal of applied physics*, 59(6):2196–2204.

CURRICULUM VITAE